लाल बहादुर शास्त्री राष्ट्रीय प्रशासन अकादमी
L.B.S. National Academy of Administration
मसूरी
MUSSOORIE
पुस्तकालय
LIBRARY

अवाप्ति संख्या
Accession No. 5870
वर्ग संख्या
Class No. 330.15
पुस्तक संख्या
Book No. Eas

330.15
Eas
5870
LBSNAA

AN INTRODUCTION TO ECONOMIC ANALYSIS

by J. K. EASTHAM

The scope of this book

is to make as comprehensive a statement of the principles of modern economic analysis as is possible at an elementary level. It was written to form the basis of the author's own teaching in economics for Part I of the new external degree of B.Sc. (Econ.) of the University of London. No previous knowledge of the subject is assumed, but the level of exposition is especially suitable for the first or second year undergraduate or the adult student who is studying economics for final profession examinations. Teachers of sixth forms taking economics at advanced or scholarship levels should find the book valuable for reference.

The plan of the book

is to take the problems of the income of a community and to construct and demonstrate the use of the tools of analysis necessary for an examination of the causes determining communal income, its distribution between social groups and fluctuations in both size and distribution. Only graphical mathematics are used in the body of the text and algebraic proofs are confined to the appendices. The problems of price determination in competitive and monopolistic markets, of productivity and of the laws of return are given a more detailed treatment than is customary at this stage, simple graphical methods being employed. The exposition of capital and interest theory is a compromise between the Keynesian and Scandinavian schools of thought, interest rates being explained as prices of loans the supplies of which are dependent on consideration of liquidity preference.

The aim of the book

is to give a coherent account of contemporary analytical technique and links with earlier theory are briefly sketched to enable the reader to connect it with a background of history of dogma to be acquired subsequently.

continued on back flap

AN INTRODUCTION TO ECONOMIC ANALYSIS

TO MY FATHER

AN INTRODUCTION TO ECONOMIC ANALYSIS

By

J. K. EASTHAM

Senior Lecturer in Economics,
Queen's College, Dundee

THE ENGLISH UNIVERSITIES PRESS LTD.

102 NEWGATE STREET

LONDON, E.C.1

First Printed 1950
Reprinted 1958

ALL RIGHTS RESERVED

PRINTED AND BOUND IN ENGLAND
FOR THE ENGLISH UNIVERSITIES PRESS LTD
BY HAZELL WATSON AND VINEY LTD, AYLESBURY

PREFACE

At some stage in the study of economics the student must acquire facility in the use of the tools of analysis if progress beyond mere description and empiricism is to be made. This step can be the first to be taken by the beginner or it may be postponed until a background of factual knowledge of affairs has been acquired. This book follows the first plan. Its purpose is to set out for the beginner in economics the first principles of the mechanism by which the world gets its living, but, since it is intended for the uninitiated, no extensive acquaintance with economic institutions can be assumed, nor can the abstract treatment appropriate in more advanced work be employed. Institutional data is therefore introduced, as required, to build up a conception of the nature of the relationships between the economic quantities which are subjected to examination.

The treatment of the subject leans to the mathematical rather than to the literary tradition in economics, but in the main body of the text there is no mathematical proposition unfamiliar to anyone who has reached a modest level of school algebra and geometry. The appendices give more rigid proofs to propositions to which it is only possible to give an air of probability by more elementary methods.

It is the purpose of the book to give as concise an account of currently established theory as is possible at an elementary level. It may therefore appear to some that certain theories which have enjoyed general acceptance until recently have received cavalier treatment. Thus the theory of demand is built up entirely in terms of the theory of choice and of substitution rates. Utility theory is introduced only in so far as it is necessary to provide a link with literature on the history of theory. Again in the theories of costs, rent and profits the method adopted is to establish directly the currently accepted theory and to show briefly its connection with earlier thought.

The study of price theory begins with the proposition of the necessary equality of marginal cost and marginal revenue for the firm in equilibrium, and from it, determines prices and outputs in various conditions of competition and monopoly. In the chapter on long-run costs emphasis is laid on a terminology which avoids the confusion arising from the indiscriminate use of the terms increasing and diminishing returns. The theory of interest adopted is the theory of the price of loans, with demand dependent on the marginal efficiency of capital, and supply on considerations of liquidity preference.

It is hoped that the book will find a place in reading lists for degree courses at whatever stage a formal study of analysis is first undertaken. Although written primarily for the student beginning his apprenticeship in economics the book is commended to the attention of the general reader who finds his grasp of analysis inadequate for his studies of economic policy.

As the book is intended for beginners, it has been kept as free as possible from footnote references. To the initiated the origins of its borrowings will be obvious, others will find the reading lists at the end of each chapter more useful. The suggestions made for more advanced reading are intended primarily for the general reader who desires guidance to more advanced work.

My debts to the work of the late J. M. Keynes and to Dr. J. R. Hicks are many. Equally onerous are those I owe to the teachers who guided my first steps in economics, the late Allyn Abbot Young and Professor Lionel Robbins. Mr. W. E. Armstrong of Southampton, and Dr. Duncan Black of Glasgow, read the manuscript in the penultimate draft and made valuable suggestions. Dr. K. C. Hayens of University College, Dundee, has saved the book from many textual obscurities and blemishes of style.

I owe particular thanks to my wife, who has taken a large share in the work of preparing the manuscript for publication.

J. K. E.

Dundee, September 1950

CONTENTS

Chapter I

WHAT IS ECONOMICS?

THE first systematic treatise of economics was called by its author *An Enquiry into the Nature and Causes of the Wealth of Nations*, and this title still serves as a broad statement of what economics attempts to do. Although the problems set out in his table of contents differ little from those in many a modern treatise, Adam Smith was under no necessity to define the boundaries of his subject. He saw himself pursuing philosophical enquiries in a direction in which other philosophers had already made tentative excursions. His was the first organised expedition into a territory about which a good deal of unco-ordinated information was available. At least from the time of Aristotle onwards men had speculated on such matters as the justifiability of interest charges and the causes determining the levels of rent and wages. The replacement of a system of subsistence agriculture by a system of exchange gave rise to new problems of taxation, and the question of the wealth of the community was of practical interest to those responsible for raising taxes.

The study of a new "subject" usually begins with the collection of data relevant to the solution of certain problems, but as time goes on investigators develop specialised techniques of dealing with certain portions of problems, and when this happens, those practising a particular technique segregate themselves from the others and define the field in which they consider they possess special competency. A familiar case is that of mediæval apothecaries whose modern descendants practise medicine, surgery, dentistry, public health, psychiatry and pharmacy, with even further subdivisions. What was once natural philosophy has given place to the natural sciences, each studying the applications of a particular specialist technique of investigation.

So it has been with economics: it does not enquire into every possible aspect of the nature and causes of the wealth of nations; it deals with those aspects which can be examined with the help of the particular technique of examination developed by economists. The full and complete study of the problems of interest, wages and rents, or of taxation, demands the use of more than one technique. There is, for instance, work for the lawyer in every one of these, but the lawyer is too busy to concern himself with those aspects of the problems which do not yield to his technique of examination, just as the economist is too occupied to attempt to acquire any one of the other techniques which might be applied when his own fails. Specialisation is necessary; the old adage that "jack of all trades is master of none" is as true in intellectual pursuits as in handicrafts.

Wealth and Welfare

To revert again to the *Wealth of Nations*, the very term wealth is capable of more than one interpretation. Smith defined the wealth of the community, in a manner very like that of modern economics, as "production per head of population." Since his time others have interpreted it differently; for instance, it has been interpreted as a state of well-being or welfare, which raises the question as to what the criteria of welfare may be. The jurist might assess the welfare of the community in terms of its legal code and the degree of observance of that code; the religious might make acceptance of certain beliefs his criterion. Unfortunately, a Liberal and a Communist would not agree in their assessments of the legal code, nor would a Roman Catholic and a Bhuddist agree on religious beliefs, and there is no absolute standard to which reference can be made to settle the dispute.

In the interval since Adam Smith, economics has endeavoured to become a science, and no science can admit questions whose answers depend on the idiosyncrasies of the enquirer, questions which are matters of opinion. Such questions may be extremely important, but they cannot be admitted to scientific discussion if the views expressed cannot be subjected to logical proof.

For a scientific treatment of any question, it is necessary to deal with quantities which can be measured, or at least it must be possible to make comparisons of orders of magnitude. We may believe that people would be "better off" if they devoted more of their constructional effort to the building of art galleries and less to cinemas; more of their agricultural output to bread and less to gin; if they printed more translations of the classics and fewer detective novels. From a moral or from an æsthetic view-point we may deplore the nature of the commodities which make up the production of a community, but we cannot argue that people are "better off" with art galleries they never visit than with cinemas which they do visit; or if the walls of their homes are decorated with prints of modernist painters they find ugly, instead of the reproductions of highland cattle which they regard as "lovely pictures"; or if they have to live on a scientifically balanced diet they find extremely dull, instead of on dietetically reprehensible menus they really enjoy.

So we must admit that our view of the state of well-being of a group of people may be quite different from their own; though the facts under examination are the same, the opinion formed on them differs with the identity of the observer. In making judgments on the welfare of a community, economics assumes that the goods and services a community consumes are selected because people prefer them to any other choice which is open to them. If they have more of the things they want, they are better off and their welfare is greater than if they had less of them. It may be the task of some branches of knowledge to explain why people choose to consume certain things in preference to others, but it is not a problem with which economics is concerned. Economics takes those choices as part of its data.

Choice between Alternatives—a Characteristic of an Economic Problem

The task of the economist, then, is concerned with the consequences of the existence of choices rather than with the reasons why people make them. He is not concerned with the reasons why white bread is preferred to brown, but with the

consequences of the choice. If only 60 per cent. of the wheat grain is turned into bread instead of 80 per cent., more wheat must be grown to supply a given quantity of bread, more land, more labour and more fertilisers and machines must be used in the production of wheat. Another corollary is that, as 40 per cent. instead of 20 per cent. is discarded, the supply of animal feeding-stuff from each ton of grain is doubled, and this in turn has its repercussions on the supply of meat and dairy products.

The choices people make reflect their desires, and these desires do not, in general, demand a unique means of satisfaction. The desire for food can be satisfied in a multitude of ways; that for warmth by expenditure on clothing, on fuel, even on food; a desire to hear good music can be gratified by expenditure on concert tickets, on a gramophone and records, or on an instrument and the tuition to enable the consumer to produce for his own needs. People may be willing to wear clothes which are exactly like those worn by everyone else, live in houses which are all identical, drive cars which differ only by their number-plates; or, on the other hand, they may demand individuality. They may be willing to buy goods by the label, pay cash and carry them away, or they may wish to compare and select, to run credit accounts and to have their purchases delivered. They may like to carry much of their wealth with them in notes or coin, or they may prefer to carry little money and use a bank account. On such choices the nature of what is produced and the manner in which it is produced depend.

Technical Facts as Economic Data

The wealth which a community can extract from its environment depends, not only on the desires of the individuals who make up the community, but on the physical possibilities of production which exist. These determine the alternatives between which choice can be made. The ease with which a supply of a particular good can be provided in a particular environment depends on a number of physical factors, such as climate and soil fertility, quantity and accessibility of mineral deposits, and on certain acquired characteristics, such as the store of accumulated knowledge and skill in the community and the

stock of artificial aids to production which it has accumulated in the past.

The production of everything we consume involves the employment of some or all of these resources. A motor-car requires so many hours of labour, so much steel, copper, tin and lead, the use of so much equipment and so much land. A bicycle requires smaller quantities of each. The production of each and every commodity requires the use of a certain quantity of the resources of the community. In a year the community can have x thousand houses and y thousand cars and z million yards of worsted suiting, and so on. Alternatively, for every twenty houses it refrains from building it can have a cinema and for every fifty it can have a factory. For two cars given up it can have a tractor. Which will the community choose ?

Now, facts of this kind are far too numerous to lie within the knowledge of any one individual. No one brain commands all the facts relating to a single commodity. Yet the alternatives exist, and choices are made between them. It is the central problem of economics to explain the process of choice.

Scarcity—a Factor in the Economic Problem

The fact that we speak of choice implies that all desires cannot be satisfied, and it is necessary to make a selection of those which are to be satisfied, leaving others unsatisfied. Some desires are of course incompatible in themselves. One cannot, for instance, enjoy simultaneously the experiences of the ascetic and the pleasures of the sybarite, but more generally the means of satisfying all desires are in limited supply. Ends are unlimited, but means are scarce.

It was fashionable in the 1930's to decry the idea of scarcity. Many physical scientists and technicians claimed that the problems of production were solved, and that, if only the system of distribution were improved, scarcity in the sense of poverty could be abolished for ever. Production here has a much more limited meaning, but even so it is not true. Even when the granaries of the world were carrying a surplus stock equal to a year's consumption, supplies were insufficient to feed the hundreds of millions of people in the world who were not eating enough food to keep them physically fit. According to

Lord Boyd Orr, the world in the period between the two world wars produced only half the food necessary to provide a *reasonable physiological minimum for the whole of its population*.

In the broader sense in which we are using the term, there is scarcity so long as the satisfaction of one desire limits the means for supplying another desire, whether of the same or another individual. So long as scarcity in this sense exists—and it is impossible to conceive conditions where it would not exist—a choice must be made between alternatives, and there is an economic problem. A commodity which is not scarce is immediately available for consumption without expenditure of time, labour or resources of any kind. Probably the only example is fresh air in the open air.

Scarce Means have Alternative Uses

Even if the supply of a particular means were limited, no economic problem would arise from its use if there were only one purpose to which it could be put. The use made of it would be purely a technical matter. If we take any article in common use, it is at once apparent that the resources which have gone to its making could have been used in other ways. The grain in a loaf of bread might have been used to make animal feeding-stuff, textile starch or alcohol; the cotton in a shirt might have been used for making a motor tyre or the wrapping of an electric cable or in the manufacture of explosive; the labour and materials which have constructed a wireless set might have been used for any one of a number of types of electrical apparatus. Selection for use in the satisfaction of one desire prevents the use of a means for some other end. When the process of adapting matter to turn it into a form suitable for human use is complete, the product is commonly so specialised as to be of little use for any other purpose, but, earlier in the process, alternatives are still open. The coal which is coked to smelt iron could have been used in a gasworks, a chemical factory or in the domestic hearth. When the steel ingot is produced, these alternatives have been ruled out, but still the steel has a wide range of uses. The labour of some steelworkers may be very specialised, but their products may still go into quite dissimilar uses.

Definition of the Economic Problem

We can now collect together what we have discovered of the nature of the economic problem. It is evidently a social *problem, and it studies the behaviour of people in the process of getting a living.* The environment in which a community lives and the powers of modifying that environment which it acquires make certain consumption experiences possible. The members of the community select, from among the consumption experiences open to them, those which they prefer, and in doing so cause the resources of the community to be directed to the satisfaction of their desires.

We can now venture on a definition of our subject :

Economics is the science of the administration of scarce means, having alternative uses, for the satisfaction of human desires.

Within this limited scope economists collect facts and arrange them in such a way as to lay bare the mechanism by which the resources of the world are organised to fulfil the needs of its population.

The position of economics among the social sciences is analogous to that of physiology among the sciences which study the human body. Physiology explores the mechanism of the human body, in so far as it can be examined by means of the techniques of physics and chemistry. Economics studies the mechanism by which the world gets its living, in so far as it can be examined by the techniques of statistics and of verbal or mathematical logic. Both sciences are concerned with the study of " what is " ; but whereas in the case of physiology " what ought to be " is a normal which can be determined by examination of a sufficiently large number of cases of " what is," in economics it is hardly possible to speak of " what ought to be." Instead we can, by a study of " what is," arrive at a series of descriptions of " what could be " ; that is, we could draw a number of designs of economic systems which might develop from that which already exists, but economic science gives no justification for singling out one of these systems and labelling it " what ought to be." The economist can, in his capacity of economist, only say, " These are the possibilities, take your choice." In his capacity of a social scientist he may go so far as to say to the community in question, " From what I know of you I think this system

will suit you best." If he goes farther and says, " This is the system you should choose," he is speaking as advocate for a particular political creed, which he is quite entitled to do provided he makes it clear that he is talking politics and not economics.

Economics and Practice

Consequently, economics " does not furnish a body of settled conclusions immediately applicable to policy," [1] either by the individual in the conduct of his own affairs or by Government authorities in the conduct of social affairs. Economics is neither the theory of business nor the theory of government. Many questions arise in both these fields to which economics can give an answer, but it cannot answer all the questions which arise in either.

The practical man of business often considers the economist incompetent to deal with the problems of his industry, because the theorist has no practical knowledge of its details. In any line of business, the difference between success and failure depends on intimate knowledge of fine detail. For instance, if one is selling cloth, it is essential to know local prejudices regarding colour, design and texture in each market in which the attempt to sell is made. The more diverse are preferences in the various markets served, the more complicated the business organisation required, the larger the stocks needed to maintain a given volume of trade and the higher the costs of production will be. The economist obviously cannot know these facts for every industry whose problems he may have to investigate, but he does know what sort of facts to look for and the effects of the existence of facts of a particular kind. The practical man is often an empiricist to whom each instance is unique, and so his problems must be worked out individually. To the theorist the method of solving a problem is to put a fact into its correct category, and if a category exists for it, theory supplies the answer to the problem.

But human problems, particularly when they are the problems of society as a whole, are rarely so simple that they can be solved by the indexing of a single fact. There are many facts and they go into different pigeon-holes, so that there is

[1] J. M. Keynes, *Introduction to Cambridge Economic Handbooks Series.*

a variety of possible solutions. If in a nation there are many people unemployed, there are many remedies which can be applied with varying degrees of success. The unemployed may be conscripted for military or labour service; they may be subsidised by the rest of the community; the State may build factories to employ them, or by subsidising agricultural wages endeavour to find more employment on the land; there is the choice between moving people to where work exists and moving work to where people are unemployed. Each remedy will not merely achieve its purpose in greater or less degree, but will also have a number of other repercussions on the life of the community. Conscription will probably be regarded as an infringement of personal liberty; subsidising unemployed may be regarded as pauperisation; a factory building programme may affect the amenities of the countryside, as will the construction of new towns if a movement of population takes place; an increase in the numbers of agricultural workers would affect the mode of life of the rural population; a general adoption of nightshift working to use equipment more intensively would affect the domestic life of the industrial population and give rise to new housing and transport problems.

Economic thinking may therefore supply a number of alternative answers to a question, and in each case the cost of the proposal and an indication of the incidental consequences would be given. A choice between the alternative policies cannot be made in terms of money alone, for other consequences, which are none the less real because they cannot be expressed in terms of money, must be taken into account. The task of interpreting the desires of the community, as they affect the choice between these alternatives, is one for the politician, not the economist.

The Method of Analysis

In this book we shall follow the example of the founder of our science and examine the causes which determine the wealth or income of a community, but we shall not expect, as he did, to discover new truths relating to the problem. Rather, our purpose will be to achieve an understanding of the nature of the tools of analysis by using them to rework questions on the answers to which economists are tolerably well agreed. It is

not the only way of approaching the subject, but it is a method which will enable us to become familiar with those tools which are most needed for the solution of the most urgent economic problems of the time. In the early part of this century economists were preoccupied with problems of the way in which prices are determined, and they found it convenient to proceed through an examination of the conditions under which the prices and quantities of all things produced would remain constant. They called their method Equilibrium Analysis, and by its aid our understanding of the pricing process was greatly advanced.

The heavy unemployment of the 1930's concentrated attention on the causes of changes in the volume of income and unemployment, and Lord Keynes' *General Theory of Employment, Interest and Money* re-established the fashion of a time when, as now, a drastic reorganisation of the economic structure of the community was in progress. What follows should be regarded as exercises in the use of a box of tools necessary for the solution of some of the problems of our time.

Further Reading

Fraser, L. M.: *Economic Thought and Language.*
Keynes, J. N.: *Scope and Method of Economic Science.*
Macfie, A.: *Economy and Value.*
Marshall, A.: *Principles of Economics*, Book I, Chaps. 1 and 2.
Pigou, A. C.: *Economics of Welfare*, Part I, Chap. 1.
Robbins, L. C.: *Nature and Significance of Economic Science.*

Chapter II

THE NATURE OF INCOME

To Adam Smith, the wealth, whose nature and causes he set out to investigate, was equivalent to the "annual produce" of the community. By this he meant its output—the quantity of goods available to be consumed directly or to be exchanged for the produce of other nations. The "annual produce" consisted only of tangible goods ready for consumption or for exchange with foreigners. He was very emphatic about this, specifically excluding services and distinguishing between those members of the community whose labour results in a tangible product and who, by their efforts, add to the produce of some future time, those whose labour "does not fix or realise itself in some particular subject or vendible commodity" but which generally perishes in the very instant of its performance, and finally, those who do not labour at all. The "annual produce" had to support all these people.

It will help us to understand the nature of the "produce" better if we think of the contributions of the various categories of people between which Smith distinguishes.

First, we have the people whose labour results in a tangible product which is immediately fit for consumption—the baker who makes bread, for instance. Then we have the man who builds a tractor, which someone will later use to cultivate land to grow wheat, which someone else will eventually turn into bread. He is contributing towards the produce of some future time, the bread supply of the year after next and some years after that. But what about the individual whose work perishes in the instant of its performance, the barber who shaves us, the domestic servant who contributes to the comfort of living, the actor whose performance we enjoy? There are many people whose services we enjoy so much that we are willing to buy them, although we thereby reduce our ability

to buy tangible goods. Why should we not include these personal services as part of the wealth of the community as well as the tangible goods it consumes? If the community were willing to enjoy fewer personal services, it could have more tangible goods—the barbers could hoe turnips and the actors cobble shoes. But the community is not willing; it makes its choice, and it would appear illogical to say its wealth is less because it secures personal services which it wants more, instead of tangible goods which it wants less.

Then let us look at those who do not labour at all. Some of them have property and live on the income they derive from it because other people use that property to produce goods. Not all of this property is situated inside the country inhabited by the "nation" or community we are considering. Such property will be used by members of other communities, and they will finally pay for the use of it by means of goods and services, which someone in our community wishes to acquire. There is also the possibility that property inside our community will be owned by people outside, and they have to be paid for the use of it. Goods and services will have to be handed over in payment.

It would appear, then, that our definition of the wealth of a community can be extended to include the tangible goods and services produced inside the community, plus those sent in from outside in payment of property owned by members of the community and used by members of other communities, minus a similar payment in respect of property inside the community owned by persons outside it.

This enlarged version of "annual produce" we will call the "real income" of the community. It is the quantity of goods and services available to the members of the community in a given time period.

Now let us look more closely at the item "tangible goods," which alone constituted wealth to Adam Smith and which still forms a major part of our definition of that quantity. We have already distinguished between goods which are ready for immediate consumption and those which will contribute to the produce of some future time. It is necessary to follow the implication of this distinction farther.

Consumers' and Producers' Goods

If we look for a moment at all the tangible goods that are made, say, within a year, we shall often find ourselves looking at the same quantity of matter in different guises at different times. Thus, looking at agriculture, we might see seed wheat and fertiliser which are put into the ground, the harvested grain, the flour and the bread. At different times during the year we shall see the same materials in these varying shapes. Or, again, in the steel industry, we might see the iron-ore, the pig-iron, the steel ingot, the forgings, castings in the rough and again when machined into finished components, and finally the finished work assembled. The finished work might be a motor-car, or it might be a lathe for turning motor-car components.

The loaf of bread and the motor-car have this in common, that they are in a state fit for direct consumption, while the wheat, the pig-iron, the forging and the completed lathe are not. The loaf and the motor-car belong to a class which are desired for consumption as they are, and in consequence are called consumers' goods. We can add to the list readily from the assortment of commodities we eat drink, wear and use daily. Some of them, notably those in the food and drink group, are capable of being used only once, and we shall refer to them as single-use goods. The others are capable of being used over a period of time, long or short, and are termed durable-consumption goods.

Other goods are not acquired to be directly consumed, but in order that some further process adapting them more closely to the needs of consumers may be performed on them, as with the wheat and the forgings; it may be they can assist in the transformation of other articles, as with the lathe. The list will comprise raw materials like coal, seed wheat and cotton, semi-finished goods like the flour and the forgings, finished components like machined castings, and finished goods forming part of the equipment of production.

Unfortunately we cannot state categorically that certain articles are always used by final consumers and so are consumption goods, and that others are always used by producers, the two groups covering the whole field. They do not. A motor-car may be used as a private means of transport, when

it will be a consumers' good, or it may be used solely to transfer people making journeys for business purposes. If a firm sends out salesmen in a car to visit customers or a general manager to visit branch establishments, the car forms part of the equipment of the firm just as much as its machinery and office equipment. We cannot therefore make our distinction in terms of kinds of goods. We must make it in terms of the uses to which goods are put.

These uses cannot be determined until goods have actually been sold to the user, so that if we wish to draw a hard-and-fast line, we shall have to adopt a definition which would make the output and sales of consumption goods always equal. For many purposes a less definite line will serve, and we talk loosely of "consumption-goods industries," meaning industries whose goods are largely purchased by private persons for their own use, but here the more rigid line will suit us better and we shall adopt it.

So far, then, we have divided the output of actual goods into two groups:

1. *Consumers' Goods*—articles ready for consumption which have passed into the hands of the consumers.
2. *Producers' Goods*—consisting of raw materials, semi-finished goods, finished goods which are in stock awaiting sale and equipment used in the production process.

Single-use or Durable Goods

Consumers' and producers' goods can each be divided into (*a*) single-use goods, such as bread, coal, petrol and many industrial stores like lubricating oil, and (*b*) durable goods, such as clothing, furniture, houses and bicycles, among consumers' goods, and all sorts of things among producers' goods which may be durable even after they have been manufactured. Many raw materials have a certain durability. Even wool and rubber can be reclaimed after they have seen service once as clothing or tyres. Metals are highly reclaimable, particularly the non-ferrous metals like copper, zinc and lead, which are reclaimed and used over and over again. The steel furnaces of Great Britain are largely fed on scrap-iron. Industrial equipment has a relatively long life; a machine tool is good for five to twelve years' hard wear; textile looms become out of

date before wearing out; buildings usually become obsolete before they must be pulled down as unsafe, and such constructions as canals, roads, embankments, artificial lakes and harbours appear almost as permanent as natural features of the countryside.

The range of durability may be anything from a few days to hundreds of years.

Services

Personal services can also be classified according to the identity of the user. Many of them are consumption goods. We go to the theatre or to concerts, have our hair cut and the like, and we are satisfying desires just as much as when we buy a packet of cigarettes. Other personal services are more like producers' goods. A modern community must devote an important part of its labour supply to administrative services, to the business of teaching and advancing knowledge. Without them the complicated machinery of twentieth-century life would not work, but we can hardly say that they result in any tangible product, although output as a whole would be much less without them than with them. For some purposes it is convenient to divide these personal services between the categories of consumers' and producers' goods, but for the moment we will keep them in a class by themselves as personal services.

Non-productive Labour

Not all writers in the past would have allowed us to include personal services in income in this way. The French economists of the eighteenth century, the physiocrats as they are now called, divided workers into two classes—productive and unproductive. The productive class contained only workers in extractive industries, mining and agriculture; people who, they imagined, were creating new stuff.

Adam Smith anticipated the results of nineteenth-century science, which denied that new matter could be created. Since atomic fission became an accomplished fact, we are not so certain about it. However, Adam Smith thought it illogical to call the smiths and weavers of his time, who were making England wealthy, unproductive workers, so he confined the

category to kings and their ministers, soldiers, opera singers, menial servants, and included himself as a university professor.

Nowadays we include as productive workers all those who by their labours contribute directly or indirectly to the satisfaction of some human need. It is true that the term is still used in more narrowly specialised senses by, for instance, cost accountants, who would include in the productive workers of a factory only those who are actually engaged in the manufacturing process, despite the fact that the work of a factory would come to a standstill within hours or days at the most if the unproductive workers did not do their jobs. In economic writing the term productive is used in its wider sense, so that there are virtually no unproductive workers.

Production as Transformation

At this point it is worth-while to state clearly what is meant by production. From what has been said already, it relates to the whole process of catering for the satisfaction of human desires. Only at the most primitive levels of human development does the natural environment of the community provide any substantial part of its means of livelihood in an immediately consumable form. In such a society, production consists of food gathering or hunting, and the making of clothing, weapons and utensils. The more advanced a society becomes, the more unpromising are the materials which it extracts in their original form from its environment, and the more complicated are the processes to which it subjects them, before they form part of an article which anyone wants to use or consume for its own sake.

Transformation of Matter

Rocks, clays and river silts are excavated and treated in an endless variety of ways until they become building materials, metals, basic chemicals, fuels and fertilisers. These are modified and combined until, after passing through a number of intermediate stages, they contribute to the emergence of something which someone wants to eat, drink or use. Often some of the original physical stuff of the basic material survives to become incorporated in the thing consumed, as when the rock mined for its metal content becomes a bicycle or the coal

becomes a dye in an article of clothing, or the clay becomes bricks in a house; but much of the original material never reaches such a shape. The iron-ore becomes a machine which makes steel tube to make bicycles, the brick goes into the factory which houses the machine, and the coal is burnt to raise steam to drive the machine. The machine is worn out in the process of making bicycles. Although the coal contributes an essential element of the process, it is not physically incorporated in the product. Nevertheless, both are just as surely transformed into bicycles as if such physical incorporation had taken place. A large part of production is thus the transformation of matter from the forms in which it is unsuited to satisfy human desires into forms in which it is better adapted to do so.

Transformations in Space

The transformation of matter is not by any means the whole of the story. Commodities in a fit state to be consumed must be moved to places where they can be consumed. Transformations in space are part of the productive process. We have heard of people dying of famine on one side of the world, while people on the other side are burning wheat as fuel because they could do nothing else with it. Without transformation in space people are dependent on the necessarily limited resources of their immediate environment, and the scope of the transformations in matter which are worth-while making is much reduced. The growth of transport services enabled wheat to be grown in the interiors of the temperate continental masses and permitted an increase in the food supply of the world.

Transformation in Time

Goods are produced at certain times, but are wanted for consumption at certain other times. Annual crops must be stored away when harvested and released gradually for consumption. Someone must store them. Someone must take the risk of loss through price changes while the goods are in store. Someone must find funds to finance the stock holding. Someone must take the risk of loss by fire or other calamity. All these services must be performed if goods are to be available when consumers

want them. We can regard them as effecting transformation in time.

Production, then, consists of transformation in matter, in space and in time.

The Causes of Wealth of a Community

When we speak of one community being rich, another poor, we often visualise the distinction as a difference in money income per head, but this is only a reflection of the fact that the annual output of goods and services per head in the one is greater than it is in the other, and we must look farther to find the cause.

The income a given community produces depends on the resources of its environment and the use it is able and willing to make of them. The resources of the environment include both its natural resources and the artificial resources which are the result of human activity.

The Natural Resources of the Environment of Society

Historically, the original base of the wealth of most communities has been soil fertility. It was apparently on the fertile alluvial soils of the great river valleys that civilisation first dawned, in those parts of the world favoured with the most genial climates, where the business of food-getting left enough leisure and energy to devote to the exploration and development of the environment. On high soil-fertility alone the early civilisations were built, but as human societies discovered fresh needs and communication between separated societies was established, other characteristics of the environment became important. Forests, which had once been mere barriers to cultivation, became important assets when timber was wanted for building dwellings or ships. Navigable rivers and natural harbours gave to those who possessed them the opportunities for activities closed to those who did not. As the use of workable stone or metals was discovered, those who possessed them could obtain means of livelihood by exchanging building stone or metal with those who did not possess them.

The geographical, geological, climatic and botanical features of a country, which comprise its natural resources, were determined long before human beings appeared on the earth, but

whether or not they are the basis of wealth depends on the ability and the desire of humans to use them. The ancients used clay for bricks and pottery, but knew nothing of the alumina it contained; nor could they have made aluminium from it had they known of its presence. Surface seepages of mineral oil may have provided fuel for fire rituals, but it was not until the invention of the internal-combustion engine that the struggle for the possession of oil deposits became a major influence in international politics.

The Acquired Characteristics of Environment

In the course of time men have modified their environment either with the deliberate intention of adapting it to their needs or by the inadvertent destruction of some of its features. Men have created artificial characteristics to serve instead of natural resources which are absent, and a good deal of constructional activity in early times was of this type. Irrigation systems were used to fertilise ground that received insufficient rain, swampy lands were drained, artificial harbours were made where natural ones were lacking, non-navigable rivers were made navigable, hillsides were terraced for cultivation, towns were built.

Roads, bridges and aqueducts, artificial lakes, land drainages and reclamation schemes, railway cuttings, embankments and tunnels are essentially artificial physical features of the environment of society, being almost as indestructible as natural features.

On the other hand, valuable mineral resources may have been partially or wholly depleted. The mineral wealth of the earth is being consumed, and in certain cases it is possible to predict the time when the deposits at present known will become exhausted. At recent rates of consumption, the alluvial deposits, which have provided most of the world's tin during the present century, will be finished in about a hundred years. At the other extreme, supplies of iron-ore seem almost inexhaustible, but, nevertheless, the expansion of steel output in this century has been possible only through the use of ores of very much lower iron content than those which were in use a hundred years ago. It is possible, of course, that technical progress may render these minerals redundant before supplies are exhausted. Alloys of the more plentiful aluminium, or even plastic material,

may be used instead, just as the synthesis of nitrates from atmospheric nitrogen has removed the danger of the exhaustion of natural nitrates.

More difficult to replace, however, are the natural soil fertilities, which have been destroyed because the consequences of methods of cultivation were not foreseen. The poisonous tooth of the goat in the Near East, the lumber mills of the North American forests, the continuous ploughing and cropping with wheat of what were once the grasslands and are now the dustbowl of North America, the constriction of the area in which the natives can practise their bush-burning cultivation in East Africa, have all had one common result—the erosion of the soil and the desiccation of the land.

Produced Means of Production

In the past two hundred years, another type of asset has become of increasing importance: goods of a durable character which help to make other goods. Tools of some degree of intricacy are probably as old as man, and devices to utilise non-human energy are of considerable antiquity, but it is only since the eighteenth century that this type of device has become so intricately elaborated and so generally used in productive processes. The community which possesses most of these mechanical aids to production can, other things being equal, produce the greatest amount of produce per head.

This effect has been described by one group of writers as the "greater productivity of roundabout methods of production." If labour is used to build equipment to make tractors, ploughs and reapers, a bigger output of grain per head is obtained than if the same total quantity of labour were used to cultivate the ground with hand-tools.

State of Technique and Skill

The construction of these valuable modifications of the physical environment and mechanical aids to production demands a certain measure of technical knowledge and skill on the part of the community which constructs them. So the powers of a community to provide itself with valuable assets are limited by the state of knowledge and technique. In a similar way there is a limit to the degree of complication in the

machinery which a particular people can use, and this in turn sets limits on the quantity and quality of the output. The considerations which determine the limits include not only the state of scientific knowledge inevitably confined to the few, but also the general " know-how " of the many and their appreciation of the limitations of mechanical devices. Before the first world war the agricultural workers of this country were more highly skilled in the manual operations of their craft than their sons and grandsons today, but their ability to handle machinery was of a very low order. Today a certain measure of mechanical skill is required in all the better-paid agricultural jobs. Many of the coloured peoples of the world in the second world war made more intimate acquaintance with the mechanical devices of the white peoples and took a first step in the process of increasing their productive power.

The ability of a given population to produce income out of a given environment depends on the state of development of scientific and technical knowledge in the community and its dispersion among the individuals comprising the community. It is dependent also on the degree of manual dexterity which the members of the community have been trained to acquire.

Willingness to Work

Ability to work is not the whole of the story. Willingness to put forth effort is an important factor also. As we shall see at a later stage, willingness to work varies between peoples and within a given people at different times. Such differences as are observed are apparently due both to differences of nurture as well as differences of nature, but it would seem most probable that differences of environment are the primary cause of the principal differences to be observed.

Forms of Organisation

Even given similar environments with similar inheritances of assets from the past, similar stores of knowledge and skill and similar degrees of ability and willingness to work, there may still be differences in income per head arising from the effectiveness of those institutions the community has chosen to construct. The form of institutions is a primary factor in determining both the degree to which human services and the nature of produced

means of production are specialised; that is, how far the community practises division of labour.

The narrower the field of tasks an individual is set to perform, the greater in general will be the degree of dexterity and skill he will achieve in them. Also, since the native ability of individuals varies in both quality and character, such a narrowing of scope gives the opportunity for specialising individuals into the tasks in which they can make the greatest contribution to the communal output.

This division of labour, as it is termed, is not confined to the specialisation of human tasks inside a community, it extends also to geographical specialisation. Certain areas can specialise in the production of a commodity to which they happen to be particularly suited. This involves an area in producing a few commodities in quantities far greater than it needs, and many others in quantities far less than it needs, so that surpluses are exchanged to make up for deficiences. Certain institutions of exchange are necessary before this can take place. Thus the degree of division of labour practised in a community and as between communities influences the volume of output.

In brief, then, the income of a community depends on the characteristics of the resources of the environment in which it finds itself and on the ability and willingness of that community to utilise those resources. Since economic thinking is concerned with the compromise between the desire of a society to modify its environment and its ability to do so, it necessarily is deeply concerned with the causes determining the size and composition of income.

Further Reading

Adam Smith: *Wealth of Nations*, Book I, Chaps. 1-3.
Cannan, E.: *Wealth*, Chap. 2.
Hicks, J. R.: *Social Framework*, Part I.
Irving Fisher: *Nature of Capital and Income*, Part II.
U.K. Annual White Paper: *National Income and Expenditure*.
U.K. Annual *Economic Survey*.

Chapter III

SUPPLY AND DEMAND

A Preliminary Survey

THE terms "supply" and "demand" are so familiar that definitions may appear superfluous, but the need for precise definition is the more urgent because they are used so commonly and ambiguously. It is one of the major difficulties for the beginner in economics, that a word which may have a variety of meanings in the language of the market must, in scientific language, be given a unique meaning perhaps not corresponding precisely to any one of its meanings in the market. This fact is a most fruitful source of misunderstanding by the lay public of the results of economic thinking, but it is rather late in the day to suggest that economists should follow the example of the natural sciences and invent a language of their own which anyone wishing to discuss their problems would first be obliged to learn.

Demand

Demand is often popularly interpreted as want. Now, for the purposes of the study of some problems, it may be possible to attribute to want a precise and definite meaning. Thus we may say that a family of a certain size wants a certain food ration, as determined by a scale laid down by some eminent physiologist; that it wants a house with a certain number of rooms, as laid down by the Housing Act, and so on. But that is not to say that the family would not wish to have more food than the minimum scale provides, or to live in a larger house, or, if it could not afford both, it might prefer to live in a house smaller than the law considers appropriate, if by so doing it could afford to purchase more than the minimum food ration. If we pass to such commodities as cinema seats, books or seaside holidays, we might find it difficult to agree on any quantity which could be said to meet the needs of a particular individual,

and, even if we could, there is no reason to suppose that we should be any nearer a solution of our problem, which is to construct an apparatus that will tell us the quantities of various commodities which will be consumed and the prices at which they will be purchased. Neither is mere desire relevant to the problem. However strong may have been the desire of an unemployed miner's family during the slump of the 1930's for new furniture to replace that which had been sold, or for a leg of mutton for Sunday's dinner, these desires had no effect on the actions of the furniture-making or sheep-rearing industries. For such an effect to be felt, desire must be backed by willingness and ability to pay a price for the desired object; thus, the number of hungry people in Manchester is not the same as the number who would be willing to pay two shillings and sixpence for a lunch. Neither can we suppose that the number willing to pay two and sixpence would be the same as the number willing to pay three shillings. There would be some, of course, who would be willing to pay the higher price rather than go hungry, but there would undoubtedly be many to whom the difference of sixpence would mean going without a midday meal if no other alternative were open. If lunches were provided only at five shillings, we should expect the number going hungry to be still greater. We cannot therefore speak of the demand for lunches, but only of the demand for lunches at a particular price.

Further, in this particular case the nature of the problem leads us to think of daily demands, but if we think of the demand for cigarettes at so much per packet or for shoes at so much per pair, we must clearly make our statement relate to a unit of time—so many per week or per year. We are now ready to summarise this in a definition thus:

The demand for a commodity at a given price is the quantity of that commodity buyers will take per unit of time.

Supply

A similar definition serves us for supply:

The supply of a commodity at a given price is that quantity which sellers are willing to offer for sale per unit of time.

The quantity offered for sale may or may not be equal to the quantity in existence. In the case of a perishable commodity,

like fish or fresh vegetables, the producer may have no alternative use for his own product, and will, in consequence, put it on the market to fetch what it will. His supply is then the total quantity he possesses. If the product is storable, the situation is different. If price falls lower than the producer considers profitable, he may prefer to store some of his product in the belief that a better price can be obtained on a later occasion. In the days before the Wheat Board, a farmer would often hold back his crop until near the next harvest in the hope of better prices. Similarly, the dairy farmer might decide to sell less milk when the price fell lower than usual, and to use the surplus for feeding pigs. At higher prices neither would be tempted to incur the trouble, expense and risk of storing grain, or of turning milk into bacon, and the product would be marketed. Except, then, in the case of a product which cannot be stored and for which the owner can find no alternative use, different supplies will be forthcoming at different prices.

Demand and Supply Schedules

It has been apparent in the above discussion that the demand for, or supply of, a commodity cannot be fully described by means of a single figure, but that it may be represented by a series of figures, each of which represents the quantity demanded or supplied at a particular price. Such a series is termed a schedule. It is no theoretical abstraction, but is an expression of facts which every successful business man must know approximately about his own commodity. Thus a successful greengrocer will know roughly how many pounds of strawberries his customers are likely to buy per week when the price is 1*s*. 6*d*. per pound, how many at 1*s*. 3*d*., how many at 1*s*., and so on. Everyday experience leads us to expect that the number sold at 1*s*. will be greater than the number sold at 1*s*. 3*d*. and the quantity at that price greater than the quantity at 1*s*. 6*d*. The increase in quantity demanded as price falls will come about in two ways: customers who may consider that they cannot afford strawberries at the higher price may be willing to purchase some at the lower, while others who were willing to purchase a small quantity at the higher price might be willing to purchase a larger quantity at the lower price. Our greengrocer would know the price at which the housewives he serves

consider strawberries cheap enough for jam-making, and the schedule would show a considerable increase in quantities at this point. The information might be tabulated like this:

DEMAND SCHEDULE FOR STRAWBERRIES

At 1*s.* 6*d.* per lb. consumers demand 40 lb. per week.
At 1*s.* 3*d.* per lb. consumers demand 50 lb. per week.
At 1*s.* per lb. consumers demand 100 lb. per week.
At 9*d.* per lb. consumers demand 200 lb. per week.

It would not be possible for a seller to form estimates over such a wide range of price for every commodity, for the seller bases his estimates on past experience of his customers' tastes, and he knows nothing of what buyers might buy at prices which are quite outside his range of experience.

In most cases the demand schedule would be of the form described, that is, smaller quantities are purchased at higher prices and larger quantities at lower prices. There may be a few cases where price and quantity are directly, instead of inversely, related; diamonds may be an example. They are likely to be cases where ostentation is a principal motive for purchase, and in consequence the higher the price the more attractive the commodity. However, it is clear that in no case could a direct relationship between price and quantity continue indefinitely. If price rises continuously, a point must eventually be reached when the number of purchasers falls off and the purchases of each of the remainder decline. Such cases can be dismissed as curious exceptions to an almost universal rule.

The Demand Curve

The use of simple geometrical methods is so invaluable in elementary analysis that we should make their acquaintance as early as possible. If we draw a graph of the demand schedule (Fig. 1), plotting prices on the vertical axis and quantities per unit of time[1] on the horizontal axis, we get what is termed a demand curve. Because, as we have already seen, large quantities correspond to low prices and small quantities to high prices, the curve slopes downward from left to right or, as the

[1] Hereafter the term *quantity* in diagrams should be interpreted to relate to a unit of time.

mathematician would say, it is of negative slope. In moving from one point to another on this curve, we are concerned solely with the effects on quantity purchased of a change in price, which we take as our independent variable. What may cause price to change does not for the moment concern us. It is usual to speak of such changes in quantity as extensions and contractions of demand, and it is said that demand extends for a fall in price and contracts for a rise in price.

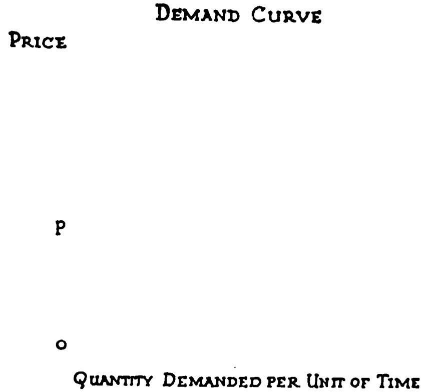


FIG. 1.

Increases and Decreases in Demand

But the quantity of a commodity purchased may vary for reasons other than a price change. Thus, the number of men's straw hats purchased today is very much less than it was forty years ago, but any change in price in the interval has probably nothing to do with it. Straw hats have gone out of fashion, men have decided that there are more comfortable and convenient forms of headgear, tastes have changed ; demand is said to have decreased. On the other hand, the quantity of cigarettes smoked today is much greater than in 1910, men have turned from pipes to cigarettes, and smoking among women, which was then unusual, is now a commonplace; tastes have changed to bring about an increase in demand.

But changes in consumers' tastes are not the only source of changes in demand. If the population of a country increases, demands for houses, for clothing, for foodstuffs and for almost every commodity or service consumed will have increased *pro rata*, provided of course that the tastes of the new members of the population are the same as those of the old members and that they are equally well able to gratify their tastes. Conversely, a decrease in population would bring about a decrease in most demands.

A further factor which can cause demands to change is a

change in purchasing power. When an individual receives an increase in income, command over goods and services in general is increased. Desires which formerly could not be gratified can now be indulged, and tastes which formerly received some small indulgence can now be gratified further. Increases in demand will not, however, be the universal result. Desires can usually be satisfied in more than one way; the pangs of hunger may be stayed by a diet of bread and cheese and potatoes or by more luxurious fare. The general increase in income among consumers of wheat during the last quarter of a century has been accompanied by a decrease in consumption of wheat per head. Cereal foods have been replaced by meat, fresh vegetables and dairy products, so that the increase in consumption of wheat has not kept pace with the increase in the number of wheat consumers. A general increase in consumers' purchasing power will therefore lead to increases in the demands for some commodities and to decreases in the demands for others.

The fact that commodities can be substituted one for another to satisfy a given desire may give rise to changes in demand in another way. If the price of one commodity rises, consumers will try to find a substitute which will serve their purpose nearly as well. Thus in 1925, when the price of raw rubber rose from 1*s.* a pound to 4*s.* 6*d.*, the demand for old motor tyres greatly increased; manufacturers thought it worth-while to incur the expense of overcoming the technical difficulties which had hitherto prevented them from using reclaimed rubber. Again, jute was practically unknown outside India until a rise in the price of flax over a century ago set sack-makers to search for a new material.

It is not necessary, however, that one commodity should be a close substitute for another in order that a change in its price should affect demand for the other. Since the consumer's income is limited, a rise in the price of one commodity may force him to curtail expenditure elsewhere. A rise in rent may mean reduction of expenditure on food. The demand for a commodity is likely to be affected by changes in the prices of other commodities, particularly of those which are substitutes.

Increases and decreases in demand, such as we have been discussing, are quite different from the extensions and con-

tractions due to changes in the price of a commodity, for they abolish the old demand schedule and the curve which represents it and replace them with new ones. If demand increases, more will usually be bought at all prices than formerly and a decrease diminishes the quantity bought at each price. Thus in Fig. 2 if the quantity purchased at price p increases from Oq to Oq_1, the change in consumers' tastes shifts the demand curve from position D to D_1, while if it decreases from Oq to Oq_2 the curve shifts from D to D_2.

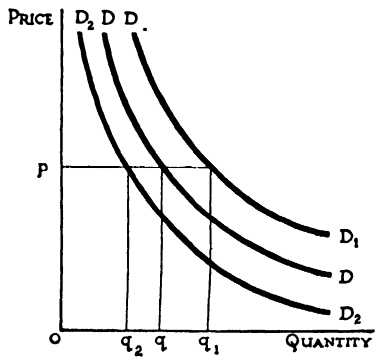


FIG. 2.

We are now in a position to state the conditions under which a demand curve will show changes in demand for given changes in price. There must be no change in (i) the tastes of consumers, (ii) the number of consumers, (iii) the purchasing power of the typical consumer, and (iv) the prices of all other commodities.

If a change in any one of these four conditions occurs, then the old demand curve disappears and a new one takes its place. A decline in the popularity of a commodity will cause a decrease in the quantity demanded at any given price, while an increase in popularity will lead to greater quantities being demanded at all prices. Changes in the number of consumers will have similar effects, but a change in the income of the typical consumer will not have a unique effect on demand.

As we have seen in the case of certain "inferior" goods, an increase in income may enable consumers to decrease consumption of such a good or to dispense with it altogether in favour of the consumption of a more expensive rival commodity.

The Interdependence of Prices

Similarly, the changes in the prices of other commodities are ambiguous in their reactions on a particular demand schedule. A fall in the prices of other commodities may lead to some of

them being substituted for the commodity under examination, in which case the demand for it would decrease. Conversely, a rise *in other prices would, in these circumstances, lead to a* demand for our commodity. On the other hand, consumers might decide that the decrease in the prices of certain commodities enabled them to buy the same quantities of these commodities for a smaller expenditure than before, and the money so saved might be spent in other directions—perhaps upon our commodity.

It might therefore appear that the factors we have outlined are changing so frequently that the idea of a demand schedule is of little importance, for the conditions underlying one particular schedule would have changed before we had had a chance to make the necessary number of observations. It is perfectly true that in strict theory a change in the price of glass marbles affects the demand for steel girders, but no one would suggest that the effect is of sufficient importance for us to take it into account in dealing with any problem of the steel market. In general, we are quite safe in ignoring all price changes except in those of commodities which are closely related to the commodity with which we happen to be dealing. Nevertheless, the fourth condition serves a most useful purpose in urging us to search carefully for connections between prices of commodities, and so complicated is the economic mechanism that this is no easy task. Furthermore, consumers as a whole are very conservative, and do not make drastic changes in their habits of consumption quickly or for small changes in price. Neither are changes in numbers likely to be sufficiently rapid to cause much trouble. Changes in the data are not likely to be so rapid as to prevent observations being made and estimates of demand to be modified accordingly.

One further word of warning may be appropriate. It is the custom in using diagrams to draw the curve from one axis to the other, but no one imagines that the data to justify this proceeding are ever known. All that can ever be justified in actuality is a small area on either side of the existing price, but as it is more convenient to draw sweeping curves than to examine diagrams through a microscope, we shall persist in the practice, but always keeping this proviso in mind.

The Supply Curve

Corresponding to the demand schedule, we have the supply schedule and its graph, the supply curve (Fig. 3). When we were discussing the shape of the demand curve, we were concerned with the tastes or desires of the body of consumers, considerations of a subjective character. Questions of tastes or desires are also relevant to an examination of the shape of a supply curve, for they help to determine the reserve price which the owner of a commodity puts on his property. To illustrate the matter, let us consider the supply curve for copies of a very scarce edition of a book. The total number in existence is fixed and approximately known, but the reserve price put upon each will not be the same in every case. Some price will be just high enough to attract an offer from the most impecunious or least appreciative possessor of a copy. At higher prices more owners will find the offer to be greater than their own valuation of their property and will be willing to sell. Finally, there is some price at which all copies which legally can be sold would be on offer. This schedule is unlike the demand schedule, in that small quantities are on offer at low prices and larger quantities at higher prices.

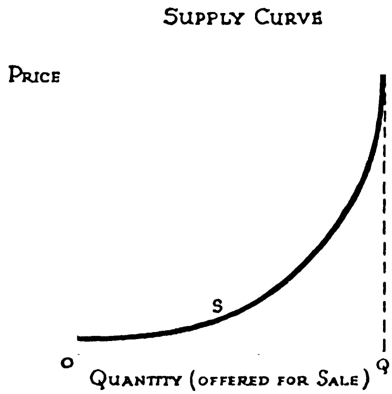


Fig. 3.

We can now draw the supply curve for a commodity the total quantity of which is fixed. The curve will begin at some point on the price axis corresponding to the minimum price necessary to attract any supply, and it will rise upward to the right, becoming steeper and steeper until, when the whole quantity in existence is on offer, it becomes vertical.

Equilibrium of Supply and Demand

We are now in a position to fit together the demand and supply pieces of our apparatus. Both pieces are as yet very

roughly shaped, and in the case of the supply piece we have taken a very highly simplified example. Our problem is to discover the price at which the market will be in equilibrium; that is, the price at which supply is such that there are no would-be buyers able and willing to pay the price but unable to find a seller, and no would-be sellers at that price unable to find a buyer. This means, of course, that supply and demand must be equal, and it follows, from the discussion earlier in this chapter, that there will be, in general, a price at which the two are equal. Our demand schedule is a scale of decreasing quantities for increasing prices, and the supply schedule one of increasing quantities for increasing prices. Thus, so long as the lowest reserve price of any seller is not greater than the highest price any buyer will pay, and the smallest quantity the sellers will offer is less than the largest quantity buyers will take, the two scales will have a common point, and in our diagram the demand and supply curves will cut. In Fig. 4 this point occurs at E, and at a price Op, Oq is both demanded and supplied. If price is lower, say, at Op_1, OA will be demanded, but the sellers will be willing to supply only OB and BA will represent the demand left unsatisfied. Would-be buyers will raise their bids and sellers will accept them, so that price will rise. Similarly, if market price is higher than the equilibrium level, say at Op_2, the quantity on offer, OC, will be greater than the quantity OD buyers will be willing to take. Sellers will undercut each other's offers, but will decrease the quantities offered and sales will increase as price falls. If, therefore, price is below Op, there is a tendency for it to rise, and if it is above Op, there is a tendency for it to fall. At Op there is no force either on the side of demand or of supply tending to change the price level, and so Op is termed the equilibrium price.

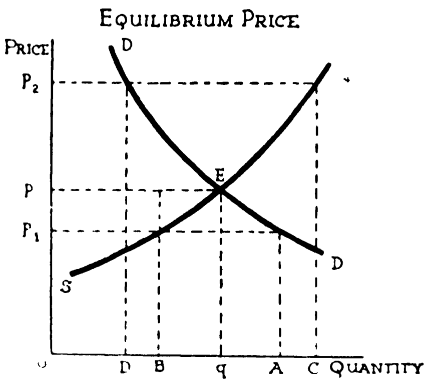


FIG. 4.

Now the statement that the market is in equilibrium at price *Op* does not mean that everybody is satisfied, in the sense that everybody who would like to buy has bought and that everybody who would like to sell has sold. It means that, given the price, neither sellers nor buyers can make a change which will better their position. All those who would like more do not demand it, because they are not willing to increase their purchases at the price *Op*, and similarly those who have the commodity would rather keep what they have than sell any more of it at the price obtainable in the market.

Changes in Conditions of Supply

A given supply curve or the corresponding supply schedule continues to represent a particular market only so long as the conditions of supply underlying it continue unchanged. Thus, if in the case of the scarce books we mentioned earlier a hitherto unknown stock was to be discovered, the supply curve would move to the right as the curve *S* in Fig. 5A has done. The conditions of supply would also be changed if the fashion of collecting scarce editions were to decline. Since the total number of copies in existence would be unaffected, the new and the old curves would coincide at the upper or right-hand end, although it is probable the curve would become vertical at a lower price, for it is likely that the whole supply would be called forth by a lower bid than formerly. Similarly, the discovery of new deposits of a mineral will move its supply curve to the right, while the exhaustion of existing deposits would cause it to move to the left, so that a smaller quantity than before would be put on the market at any given price. In the case of manufactured articles, the invention of a new process may enable any given quantity to be produced at a lower price than before, so that, at a given price, the quantity supplied is greater than before.

Equilibrium with Changes in Demand and Supply

We can now consider the movements in price which are likely to take place when changes in both the conditions of demand and supply occur. We start from the equilibrium position where price *p* is determined as in Fig. 4 by the intersection of the demand curve *DD* and the supply curve *SS* (Fig. 5A).

First, we will suppose that the commodity becomes more popular, so that more of it is purchased at all prices and the demand curve moves into the position D_1D_1. Nothing has as yet happened to the conditions of supply, so that price will now be determined by the intersection of the new demand curve with the old supply curve, that is, it will be Op_1. But the increase in demand for this commodity and the rise in its price

CHANGE IN CONDITIONS OF DEMAND AND SUPPLY

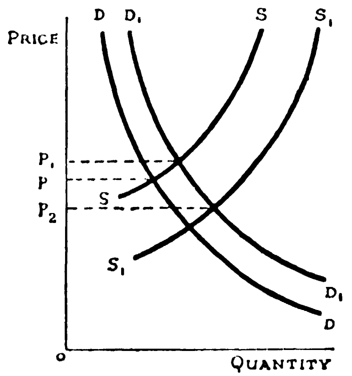


FIG. 5A.

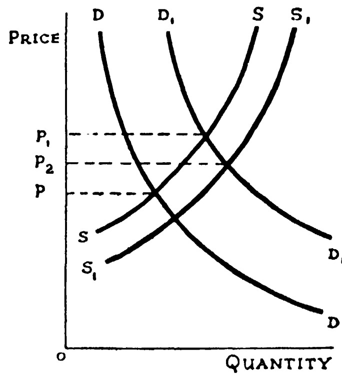


FIG. 5B.

may induce other people to manufacture it, so that more will be supplied at all prices and the supply curve will move into the position S_1S_1. Price will now be determined by the intersection of the new supply curve and the new demand curve, i.e. at Op_2. The final price Op_2 will certainly be lower than Op_1, but how it compares with the original price Op depends on the relative shifts of the demand curves. In Fig. 5A, where the shift in the supply curve is relatively greater than that in the demand curve, price is lower when the changes are completed; while in Fig. 5B the shift in the demand curve is the greater and price has risen. Equal shifts would, of course, leave the price unchanged.

Such changes, particularly those in supply, take time to accomplish, and the movement from one curve to another is a gradual process, while movements up and down a particular demand or supply curve are considered to take a negligible time. Indeed, to be quite exact we must regard points on one

of these curves as representing alternatives, which are open at any moment, given the conditions of demand or supply. It is probable that the case illustrated in Fig. 5A is the one most frequently to be met in practice. A random selection of the inventions of the present century will provide numerous examples of a great increase in demand accompanied by a fall in price. A spectacular case is that of rubber. In 1910 the world consumed 88,000 tons of rubber, at an average price of 8*s.* 9*d.* per pound, while in 1930 it consumed 710,000 tons at a price of about 5*d.* per pound.

Price and Cost

It may have occurred to some to enquire whether cost of production has anything to do with price. We dealt first with a simplified case where a fixed quantity of a commodity was assumed to exist. In this case any original cost of production is quite irrelevant. Price depends solely on the tastes of those who possess and those who wish to acquire. Indeed, it is possible to go farther, and to say that once any commodity is in existence its cost of production is quite irrelevant to its price. A supply is in existence, and must be sold for what it will fetch in the market or not at all. But although the cost of production cannot affect today's supply, yet the anticipated cost of reproduction does affect tomorrow's supply. Unless market price is such that it covers the cost of reproducing today's supply, then production for tomorrow will be curtailed or may cease entirely.

In this chapter we have constructed a crude working model of the mechanism of price determination. The conditions underlying the demand and supply curves will be further investigated and the model elaborated in subsequent chapters. In conclusion we will set out the generalisations which have emerged from this discussion:

1. If price rises, other things being equal, demand contracts; conversely, if price falls, other things being equal, demand extends.
2. If price rises, other things being equal, supply extends; conversely, if price falls, other things being equal, supply contracts.

3. Price tends to the level at which supply and demand are equal.
4. If demand increases, conditions of supply remaining the same, price rises; conversely, if demand decreases, price falls.
5. If supply increases, conditions of demand remaining the same, price falls; conversely, if supply decreases, price rises.

Further Reading

Boulding, K. E.: *Economic Analysis*, Chaps. 3 and 4.
Henderson, H. D.: *Supply and Demand*, Chap. 2.
Wicksteed, P. H.: *Commonsense of Political Economy*, Book I, Chap. 1.

Chapter IV

THE THEORY OF CHOICE

In the previous chapter we arrived empirically at certain generalisations on the relationship between the price of the commodity and the quantity of it demanded. We reached such broad conclusions as that which states that the quantity demanded is larger for a lower price than for a higher price, and that changes in certain fundamental conditions will cause the quantities demanded at all prices to increase or decrease. We did not find any explanation why these movements should be of greater magnitude in certain cases than in others, or any indication of the underlying mechanism. It is this deficiency we must now repair.

So far, in the discussion of the behaviour of people disposing of a fixed income, we have kept the argument at the level of deduction from common experience, and we shall maintain it there.

The problem we have now to solve is to determine the relationship between the price at which a commodity is sold and the quantity of it a given individual will purchase at that price. The traditional approach to this problem was by consideration of the satisfaction an individual was supposed to derive from the consumption of successive units of the commodity. The method involves the use of psychological assumptions of dubious validity, which are not essential to the solution of the problem. Therefore, on grounds of the well-known logical principle of economy of hypothesis, an approach to the problem which involves fewer and less-complicated hypotheses should be preferred.

We will approach the matter from the standpoint of an individual dividing his expenditure among the various rival claims against it. He has a certain amount of money income and he buys certain goods and services in the market and

possibly sets aside a certain amount of cash as reserve, but for the moment we will suppose that all private persons, as distinct from persons organising businesses (though these will sometimes be the same persons acting in another capacity), spend all their income on goods and services for consumption.

It would be quite unreal to picture an individual planning what he was going to spend on each and every one of the goods and services he wished to consume, yet most individuals have habitual levels of expenditure, and are accustomed to consume reasonably constant quantities of each. These levels are arrived at by a process of trial and error, and are subject to some adjustment from time to time. The adjustment comes about through the individual reviewing expenditure and deciding whether it could not be better distributed. There is a constant tendency to cut down expenditure on some lines in order to spend more somewhere else.

Comparisons are made of the modifications which are possible. Cutting down the number of cigarettes smoked by twenty a week will permit the purchase of a concert ticket once a fortnight, or will enable a slight improvement in lunches to be made twice a week. These alternatives are set one against the other and a decision is made. We need not beg any psychological questions as to whether people really prefer the alternatives they choose to those they reject. We are concerned solely with their behaviour, and if, when faced with the choice of A or B, they take A, we shall say they prefer A. By the term preference we shall mean merely that the persons concerned are in the habit of making one choice rather than another, when both are open to them. If, therefore, there happens to be an unallocated shilling and the decision is taken to spend it on oranges, this will constitute proof that the consequent additional oranges are the most important contribution to consumption which could have been made with the shilling. If the shilling has been made available by a reduction in the expenditure on potatoes, we shall take it as evidence that, at current market prices, a shilling's-worth of potatoes is considered as of less importance than a shilling's-worth of oranges.

If, this adjustment having been made, there is no further deduction from consumption of potatoes in favour of the consumption of oranges, we are to conclude that no advantage lies

in a further exchange. Given the quantities of each of the two commodities which are now being consumed per week, there is no advantage to be gained by further substitution of one for the other. An additional shilling's-worth of oranges and the shilling's-worth of potatoes which would have to be surrendered to get it are now considered as equally valued. If a shilling buys a stone of potatoes or six oranges, we have thus determined the valuation of the one in terms of the other for the individual in question, when certain quantities are being enjoyed per week.

Marginal Rate of Substitution

Now it has already become apparent that this valuation of small increments of one commodity in terms of another varies with the total quantities of the two the individual concerned possesses. The last stone of potatoes was clearly worth less than the six oranges which replaced it in consumption, but this increase of one quantity and decrease of the other modified valuations, so that after the substitution was made, six oranges were no longer worth more than a stone of potatoes. It is useful to have a term to indicate this valuation of one commodity in terms of another, and we shall call it the Marginal Rate of Substitution (M.R.S.). We will define it as the ratio between small quantities of two commodities which are equally valued.

Thus, if an individual considers that, by increasing his consumption of tea by one pound a month at the expense of cutting down his tobacco consumption by an ounce, he would be no better and no worse off, then we can say that the marginal rate of substitution of tea in terms of tobacco is $\frac{1 \text{ oz.}}{1 \text{ lb.}}$. It is not suggested that the comparison would be made in physical terms in this way, but the actual process would come to the same thing. The cost of the ounce of tobacco and of the pound of tea is the same, and the consumer considers saving on the one in order to spend on the other. In this case we can get a measurement of the marginal rate of substitution in physical terms, since the increments of the two commodities are expressed in the same units and 16 oz. of tea are considered equal to 1 oz. of tobacco. Thus $\frac{1}{16}$ is also the inverse ratio of the

prices of the two commodities. We can say, therefore, that when an individual has found an equilibrium between consumption of two commodities, the marginal rate of substitution of one in terms of the other will be equal to the inverse ratio of their two prices.

Personal Valuations and Market Valuations

Each individual who is balancing expenditure has therefore two valuations to bear in mind: (*a*) his own personal valuation of an increase in the consumption of one commodity made at the expense of a decrease in the consumption of the other, i.e. his marginal rate of substitution, and (*b*) the market rate of valuation or price. Let us suppose that we are concerned with an individual who consumes only two commodities, say bread and butter. If the market valuation of butter is not so great, relatively, to that of bread as our individual's personal valuation, he will be willing to give up bread in exchange for butter, and will continue to do so as long as this state of affairs continues to exist. Exchange will come to an end if and when the private valuation and the market valuation are equal, that is, when the marginal rate of substitution of one for the other is equal to the inverse ratio of their prices. When he has reached this point, he will be unwilling to increase his consumption of either commodity at the expense of the other, because in each case the increment of one will not be worth the small quantity of the other which must be given up to obtain it.

Similarly, if a person's valuation of butter in terms of bread is less than the market valuation, he will be willing to give up butter and buy bread, and again will continue to do so until either the market valuation or his own valuation changes sufficiently for the two to reach equality. We have therefore the following condition of equilibrium of exchange for two commodities and money:

The marginal rate of substitution of A for B

$$= \frac{\text{A small quantity of } B}{\text{The small quantity of } A \text{ equivalent in estimation of individual to above quantity of } B.}$$

$$= \frac{\text{Price of } A}{\text{Price of } B}.$$

Diminishing Marginal Rate of Substitution

The argument we have used to determine the conditions of market equilibrium presupposes that the marginal rate of substitution of one commodity for another decreases as the quantity of the commodity possessed increases. Thus, if an individual is giving up *B* and acquiring *A*, it is necessary to the argument that he should demand more and more of *A* in exchange for a given quantity of *B*, or, to put it another way, he is willing to give less and less of *B* for successive equal increments of *A*.

Is it really necessary that this condition should hold, and what would be the consequences if it did not hold?

If the marginal rate of substitution of one individual for one commodity in terms of another should remain constant at a level higher than price irrespective of the quantity acquired, exchange would continue until the whole of income was spent on the one commodity. As we are assuming that we are dealing with an economy in which there are only two commodities, the marginal rates of substitution of each in terms of the other are reciprocals, so that if the M.R.S. of *A* for *B* is constantly above the price ratio, the M.R.S. of *B* for *A* must be constantly below the price ratio. In these circumstances an individual would sell all the *B* he possessed and spend all his money on *A*.

Such a situation must be admitted to be possible, but if quantities of both commodities are consumed, we must have the condition of decreasing M.R.S. if we are to explain the relative sizes of the two quantities.

What condition would account for this phenomenon of a decreasing marginal rate of substitution? It would occur if, for at least one of the commodities, it were possible to arrange uses for the commodity in descending order of importance. Suppose we consider the butter supply consumed by an individual to be divided into ounce lots and that specific uses can be found for successive ounces. The first ounce is devoted to the use considered most important by the consumer, and the second to the next in order of importance and so on.

If this were so, the M.R.S. of butter for bread would decline as more butter was acquired, even if the uses of successive loaves of bread were considered equally important. If at the

same time successive uses for a loaf of bread were considered to be of decreasing importance by the individual in question, the M.R.S. would decrease all the more rapidly. Not only is each ounce of butter acquired meeting a less important use than the one before it, but each successive loaf given up is being subtracted from a more important use than the one before it. It is not an unreasonable generalisation from experience that, for most commodities, uses can be arranged in descending order of importance.

Exchange with More than Two Commodities

We can apply the rule we have discovered for the exchange of two commodities to a market in which any number are exchanged. Between each pair of them we must have the same relationship; the marginal rate of substitution will be equal to the price of one in terms of the other.

If ΔA, ΔB, ΔC, etc., are small quantities of A, B, C, etc., which are equally valued, and $p_{A/B}$ signifies the price of A in terms of B:

$$\frac{\Delta B}{\Delta A} = p_{A/B}; \quad \frac{\Delta C}{\Delta B} = p_{B/C}; \quad \frac{\Delta D}{\Delta C} = p_{C/D}; \text{ etc.}$$

Again, this involves nothing more than the assumption that an individual experiments with consumption and increases the quantity of one good consumed and decreases consumption of another, until he arrives at the ideal collection of goods and services purchasable with his income.

But when we take this step to a multi-commodity market, the use of prices which are exchange ratios between any two commodities—the price of A in terms of B and the price of B in terms of C, etc.—becomes very cumbersome, and it is more convenient to use one of the commodities as a standard. Then all prices are measured in terms of the standard, which we will suppose is wanted only because it can be used for making payments and for no other purpose. The standard has then taken on one of the attributes of money, and we shall now call it money, remembering that until further notice it is money only in a limited sense of the word.

We can now use an alternative form of the argument above. Because the equilibrium condition above holds between each

pair of commodities, it holds between each of them and money, so the marginal rate of substitution of each commodity for money will be equal to its money price; so that, if ΔM, ΔA, ΔB, etc., are equally valued small quantities of commodities, M being used as money:

$$\frac{\Delta M}{\Delta A} = p_A; \ \frac{\Delta M}{\Delta B} = p_B; \ \frac{\Delta M}{\Delta C} = p_C; \text{ etc.}$$

Now, we can get this result another way. When an individual has distributed his income among the various objects of expenditure open to him, he will be unable with advantage to switch expenditure from one commodity to another. Therefore, if a small quantity of money ΔM buys a small quantity ΔA of commodity A, or a small quantity ΔB of B, etc.:

$$\Delta M = \Delta A.p_A = \Delta B.p_B = \Delta C.p_C, \text{ etc.}$$

So again we have:

$$\frac{\Delta M}{\Delta A} = p_A; \ \frac{\Delta M}{\Delta B} = p_B; \ \frac{\Delta M}{\Delta C} = p_C; \text{ etc.}$$

And again we have the marginal rate of substitution of the commodity for money equal to its money price. When this condition is satisfied, an individual cannot increase the quantity of any commodity consumed per unit of time at the expense of any other commodity with advantage to himself.

We have now arrived at a most useful piece of information. We know that *the quantity of a given commodity an individual will consume at a particular price is that which makes his marginal rate of substitution for money equal to the price.*

Diminishing Marginal Rate of Substitution Reconsidered

It can now be shown that a constant marginal rate of substitution is much less probable than it appeared when we discussed only two commodities. If the marginal rate of substitution of a commodity for money remains constant, an individual will either buy none or spend the whole of his income on it. He will buy none if the M.R.S. is constant and less than the market price. He will spend all his income on it if the M.R.S. is constant and greater than price. If by coincidence the M.R.S. and price happen to be equal, expenditure might increase to total income, or decrease to zero, in response to any accidental change.

It is not necessary to assume, as the utility theorists did, that in every case successive increments are put to uses of decreasing importance. Even if in some cases the uses for a commodity are of constant importance, as more of it is acquired the M.R.S. will still diminish so long as the condition of constant importance of uses does not hold for all other commodities. The possibility of a preference scale for a commodity where successive uses are of increasing importance can be ignored. All uses are supposed to be alternatives at a moment of time and, if consumption fosters appetite, uses have to be taken in order of time. We are then concerned with a preference scale which is changing rapidly, and not with movements on one scale.

M.R.S. and the Demand Schedule

We are now in a position to examine much more closely the nature of the demand schedule, whose acquaintance we made in Chapter III. Our study of the marginal rate of substitution has shown us that the quantity an individual is willing to consume at a given price is that quantity which makes his M.R.S. equal to the price. So if an individual starts buying a commodity *A* at price p, he will work his way down his preference scale, filling successive uses, until he comes to the point where the satisfying of an additional use will leave him no better off, because the uses of other things he will have to sacrifice are just as important to him as the use for *A*, whose fulfilment he is contemplating. His M.R.S. is then equal to the market price p_1 and the quantity consumed per unit of time is q_1. If price falls, say, to p_2, he will have to increase his consumption by some amount, great or small, before the condition of equilibrium is again satisfied. So we can proceed for each member of a series of prices.

We can go through the same process with all the individuals concerned in turn. Some individuals may start buying at prices higher than p_1, others may buy nothing until price has fallen below this level. At the end of our investigation we shall have a scale of prices, at each of which at least one individual is willing to consume some quantity of *A*. For each individual, the lower the price the larger the quantity consumed; and for the market as a whole, the lower the price,

the more individuals are consuming at that price, until all are consuming. We can therefore add up the quantities that individuals are willing to buy at any particular price, and this *constitutes the market demand at that price.*

Thus in Table 1 we see the construction of a demand schedule for cigarettes for a market consisting of three people X, Y and Z. At a price of 3*s.* 6*d.* per packet, X buys two packets, which quantity equates his M.R.S. to the market price. For Y and Z, price exceeds the M.R.S. for any quantity, however small. At 3*s.* X buys three packets per week, because when he is doing so his M.R.S. is again equal to the new price. Y now buys one packet, as when he does so his M.R.S. is also equal to 3*s.* but not until price has fallen to 2*s.* 6*d.* does it become equal to the highest value of M.R.S. on Z's scale and he begins to consume. The total demand for the market is obtained by adding the quantities individuals will purchase at each price, and the idea can be extended for a market consisting of any number of individuals.

TABLE 1

DEMAND SCHEDULE FOR CIGARETTES

Price	*No. of Packets purchased per Week.*			
	X	*Y*	*Z*	*Total*
3*s.* 6*d.* .	2	0	0	2
3*s.* . .	3	1	0	4
2*s.* 6*d.* .	6	2	1	9
2*s.* . .	8	3	2	13
1*s.* 6*d.* .	9	5	4	18

This method of deriving a demand schedule for a commodity brings out certain features which other methods obscure. It is quite apparent that the demand schedule for one commodity is not dependent alone on the tastes of consumers for that commodity. The quantity of tobacco consumed does not depend solely on consumers' desire to smoke, but also on their willingness to divert income from other objects of expenditure. So a given demand schedule depends, not only on tastes for the commodity to which it relates, but also on tastes for all other commodities which consumers purchase or contemplate purchasing.

If consumers' tastes for sweetmeats change, their valuation of tobacco in terms of sweets changes, but market prices have

remained the same. Therefore, the quantity of sweets which can be bought for a shilling, which formerly was valued equally with the quantity of tobacco a shilling will buy, is now more highly valued. In order that the balance shall be restored, consumption of sweets has to be increased and consumption of something, perhaps tobacco, has to be reduced. Similarly, if the price of sweets falls, consumption of some other commodity will probably be adjusted to restore the equilibrium. The prices of all consumption goods and the quantities of them demanded are thus interdependent, but to trace their interactions we must examine some of the relationships between commodities.

Rival Goods

The most common relationship between consumers' goods is that of rivalry. Finally, this is the relationship between all commodities and services, as all are rivals for a fixed total of expenditure. Many commodities are, by nature, in some degree substitutes for each other. Thus, all forms of food are in some degree rival, as one form can be substituted for another, and we have already seen that the possibilities of substitution do not stop here. The more generalised desires may be, the more diverse the range of goods and services which can cater for them and which in consequence are substitutes for each other.

Thus, if the price of mutton falls so that an individual moves farther down his scale of diminishing marginal rate of substitution of mutton for money, his desire for beef and hence his personal valuation of beef will be affected. It will not merely be a matter of his valuation of the last incremental unit of beef in the least important use to which he put it, but his whole scale of valuations of beef in all its uses will tend to fall as he substitutes mutton for beef.

The effect of a fall in the price of one of a group of rival commodities will therefore be to depress the scales of marginal rates of substitution for money of the other members of the group

Complementary Goods

In contrast to the rivalry relationship between commodities, it is possible that there may be a relationship between two or more commodities such that, if for any reason whatsoever the

consumption of one of them increases, desire for the other becomes stronger and personal valuation of it is higher. In such a case, a fall in the price of one commodity causes an individual to travel farther down his scale of diminishing marginal rate of substitution of the commodity for money, and in doing so to consume more of the commodity. This results in a raising of the scale of marginal rates for the related commodity, so that, if its price remains the same, more of it will be consumed in order to descend the scale of marginal rates to the point where the marginal rate of substitution for money is equal to the money price once more.

The simplest cases of such a relationship are those between meat and mustard, bread and butter, tea and milk, fresh fruit and jam sugar. In the case of motor-cars and petrol, the effect for a single individual is to be noted only at the point where the fall in the price of cars is sufficient to induce a non-owner of a car to become a car owner, but as regards the total demand for cars, the principle is still operative. An increase in the number of cars purchased would lead to an increased demand for petrol.

Income Effect and Substitution Effect

If a housewife has been in the habit of buying 200 lb. of butter per year at 1*s.* 8*d.* per pound and the price of butter falls to 1*s.* 6*d.* per pound, the position is somewhat analogous to that of an increase in income. There will be, as we have seen, a tendency to substitute butter for other things, simply because further units can be obtained at a lower price. But the situation would be no different if the housewife were given an increase in income of 400*d.* per annum and the right to buy 200 lb. at 1*s.* 8*d.*, and as many more as she liked at 1*s.* 6*d.* The adjustment which would then take place would be as between the extra income and the right to purchase more butter at the lower price. The extra income may be spent wholly or mainly in buying more butter, or expenditure may be diverted to the purchase of other commodities. Thus the effect of a fall in the price of a commodity can be broken up into two distinct effects. The first we have already noted. A commodity whose price has fallen is substituted in use for other commodities which are more or less closely rival to it. This phenomenon we term

the substitution effect, and the expenditure for the increase in consumption is found from a decrease in expenditure on the commodities which are substituted against.

The second effect is concerned with the disposal of the release of expenditure which arises because it is possible to buy the same quantity of the commodity whose price has fallen for a smaller expenditure than before. This is termed the income effect. Taken alone, it may lead either to a maintenance of expenditure on the commodity whose price has fallen or to a decrease in that expenditure.

Inferior Goods

In one type of case, the tendency to divert expenditure away from a commodity whose price has fallen may be so strong that, not only is the saving on the reduced price spent on something else, but consumption of the cheapened commodity may actually be reduced. For this to happen it is necessary that either the marginal rate of substitution of the other commodity is raised by the price fall, or the scale of marginal rates for the cheapened commodity is reduced. This occurs when two commodities technically fulfil substantially the same purpose but one is considered superior to the other.

The cheapest common foodstuff is usually in this category. People who consider that wheat flour is a choicer article of diet than potatoes nevertheless may have to buy potatoes exclusively because, given their income and the ruling price of potatoes, they cannot provide an adequate diet when buying both. A fall in the price of potatoes may enable them to reduce potato consumption and to substitute the more expensive wheat flour. In these circumstances the income effect of a reduction in the price of potatoes is said to be negative.

Further Reading

Fetter, F. A.: *Principles of Economics*, Chap. 2.
Knight, F. H.: *Risk, Uncertainty and Profit*, Chap. 3.
Phelps-Brown, W.: *Framework of the Pricing System*, Chap. 2.
Wicksteed, P. H.: *Commonsense of Political Economy*, Book I, Chaps. 2 and 3.

Chapter V

INDIFFERENCE CURVE ANALYSIS

THIS chapter presents a more precise method of arriving at demand schedules for consumption goods than that offered by the Theory of Choice in Chapter IV. The mathematics used involves nothing more profound than a little elementary co-ordinate geometry, but it permits a far clearer statement of the principles of pricing than is possible in purely literary terms.

The Indifference Curve

The method of scales of marginal rates of substitution we have been using involves keeping in mind the personal valuations of individuals, their incomes, the price of the commodity under examination, and the prices of other commodities which enter into consumption, but it does not offer any visual apparatus to assist us. A slightly different approach to the matter still retaining the idea that the marginal rate of substitution measures the personal valuation of the individual, is sometimes found simpler by beginners, and in more advanced studies provides an alternative to somewhat complicated algebraical methods.

Let us return to the individual whom, in the previous chapter, we studied adjusting his consumption as between bread and butter. We will suppose he has x units of bread, y units of butter and nothing else whatever. These are the only commodities entering into exchange. Now it may be that if our individual had x_1 of bread and y_1 of butter, instead of x of bread and y of butter, he would consider himself equally well off. So we might get a whole set of pairs of quantities of the two commodities, which would be equally welcome to the person in

question. These we could plot in a curve, I, Fig. 6A, whose horizontal axis is quantities of bread and whose vertical axis represents quantities of butter. It is the locus of the points

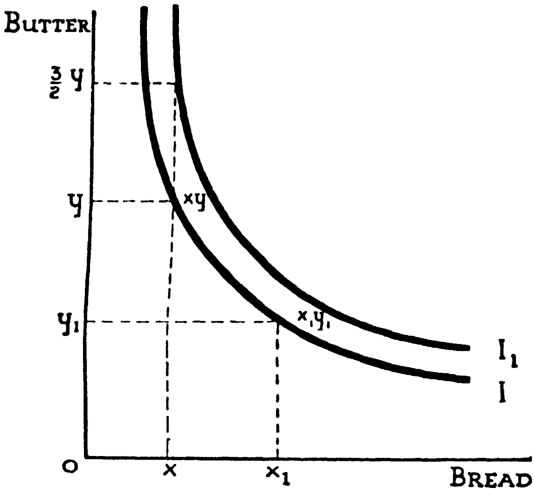


FIG. 6A.

representing pairs of quantities between which the individual is indifferent, so it is termed an indifference curve.

We can then draw another like it, starting with x units of bread and $\frac{3y}{2}$ units of butter, and draw the locus of points representing combinations between which the individual is again indifferent. In this way we can derive a whole family of curves above and below the first one at whatever intervals we choose.

Now let us return to the curve which shows pairs indifferent with x and y. The point $x_1 y_1$ is on the curve close to xy, and could be reached by an individual at xy giving up $y - y_1$ of butter and getting $x_1 - x$ of bread in exchange for it. These small doses of each of the two commodities are equally valued by the individual to whom the curve relates, as he is indifferent whether he is at xy or $x_1 y_1$. Therefore:

$$(y - y_1) \,/\, (x_1 - x)$$

is the marginal rate of substitution of bread in terms of butter

at this point, and this is represented geometrically by the slope of a straight line joining the points xy and x_1y_1.

If we take the two points close enough together, the chord to the curve and the arc of the curve become identical and can both be represented by the slope of the curve. *The marginal rate of substitution between the two commodities, when any given combination of them is possessed, is the slope of the indifference curve passing through the point representing that combination.* Between the two axes we can find a point which will represent a combination of any given quantities of the two commodities which we may choose, and the marginal rate of substitution corresponding to that combination can then be discovered by constructing an indifference curve to pass through that point.

Diminishing Rate of Substitution and the Shape of the Indifference Curve

An application of the arguments we have used to demonstrate the principle of diminishing marginal rate of substitution can be used to determine the shape of the indifference curve. We saw that in general the marginal rate of substitution decreases as more of one commodity is acquired at the expense of another. It is necessary therefore that the quantity $(y_1 - y)/(x - x_1)$, which represents the marginal rate of substitution of bread in terms of butter, shall decrease as the quantity of bread is increased. This means that the indifference curve must slope downwards from left to right and must be concave upwards. As we saw before, we cannot be very sure of the truth of this proposition when we are dealing with two commodities only, but more reliance can be placed on the proposition when a number of commodities enter into exchange.

Prices on the Indifference Map

Just as the variable personal valuation M.R.S. can be represented on the indifference diagram by the slope of the indifference curve, so a price can be represented by the slope of a straight line. If *OL, OM* and *OR, OS* (Fig. 6B) are two pairs of quantities of bread and butter, and an individual can, by exchanging one commodity for the other, have either of the two combinations, then the rate of exchange between the two, or the price of bread in terms of butter, is the quantity of one

which has to be given up in the market divided by the quantity of the other acquired, that is:

$$\frac{OM - OS}{OR - OL} = \frac{SM}{LR} = \frac{PT}{TP_1}$$

which is the tangent of the angle the line PP_1 makes with the bread axis.

Conversely, if it were butter which was being acquired, the price would be represented by the inverse of this quantity or:

$$\frac{OR - OL}{OM - OS} = \frac{LR}{SM} = \frac{TP_1}{PT}$$

the slope of the line PP_1 measured from the butter axis. So we can represent price by the slope of a line drawn through the

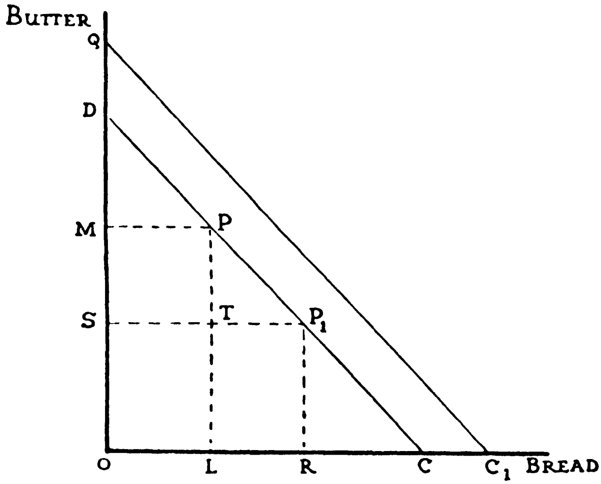


FIG. 6B.

point representing the quantities of the two commodities the individual happens to possess. Such a line also represents the locus of points representing all the combinations of the two commodities which the individual can reach at the given market price. It is the path along which exchange can take place if the individual in question so desires. How far he will choose to travel along it we shall see presently.

All the price lines drawn through any point will slope downward from left to right. This merely reflects the fact that it is

always necessary to give up something in order to acquire something in exchange. If a price line is parallel to either axis, the commodity represented by that axis is a free good. If the line slopes upward from left to right, price is negative: the buyer is paid for taking it away.

Income on the Indifference Map

We have already shown implicitly how the income of an individual may be shown on an indifference map. We have seen that the quantities of the two commodities an individual possesses before he starts to exchange can be represented by a point on the indifference map. In dealing with the problem of the division of a given income between competing objects of expenditure, it is simpler to start with the assumption that an individual receives his income in the form of a quantity of one commodity and exchanges some of it for the other commodity. Income is then represented by a point on one or other of the axes, its distance from the origin representing the size of income in terms of one commodity alone. Thus, in Fig. 6B, OC represents a given income measured in terms of bread and OC_1 a bigger income measured in similar terms. A straight line CD, drawn from C, represents by its slope a given price of bread in terms of butter, and, as C_1Q is of similar slope, it represents the same relative price of the two commodities.

But these lines represent something else as well. If an individual starts off with OC of bread and exchanges some of it for butter, his income remains the same in total value, though it consists of two commodities instead of one. Thus the line CD represents the pairs of quantities of bread and butter it is possible to buy with an income equal to either OC of bread or OD of butter, when the price of bread in terms of butter is OD/OC. The line CD is therefore a constant income line and C_1Q represents another constant but larger income.

Determination of Quantity Purchased, given Income and Price

In Fig. 7 we are concerned with an individual who has an income of OY units of commodity B, but who wishes to acquire commodity A. His personal valuations, in all possible income circumstances, are represented by his system of indifference

curves. Price can be represented by any line of appropriate slope. We will suppose that one unit of A is valued by the market at $\frac{9}{8}$ units of B, then this market price can be represented by any straight line making an angle $\tan^{-1}\frac{9}{8}$ with the A axis. If we draw such a line through point Y, it will represent all the points which an individual can reach from the starting-point set by his income, so long as price remains constant at this level. Imagine, then, an individual setting out from point Y and exchanging B for A at the rate of one and one-eighth units for one. At a he encounters an indifference curve, but this is no stopping-place. In his opinion it is worth giving up ab of B in order to get bc of A in exchange, but the market will give him a larger quantity than this, i.e. bd.

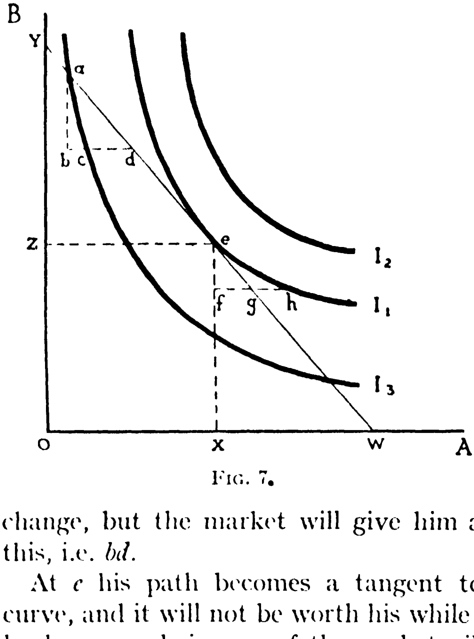


FIG. 7.

At e his path becomes a tangent to another indifference curve, and it will not be worth his while to pass this point. If he does so and gives up ef, the market will give him in exchange fg, but in his opinion ef is worth fh, a larger quantity. He will therefore go no farther than point e. He will have acquired OX of A, and given up YZ of B in exchange, and will be in the position of having divided his income between the two commodities available to him. In this case he has given up a number of units of B equal to $\frac{9}{8}$ times the number of units of A he has acquired, but this is merely the result of the particular market price we selected. If we had selected a higher price for A, it would have been represented by a line falling more steeply than YW, and equilibrium might have been found at a point on a lower indifference curve. Similarly, had the price of A been lower, it would have been represented by

a line falling less rapidly than YW and equilibrium might have been found on a higher indifference curve.

Exchange in a Multi-commodity Market

The process we have just worked through for a pair of commodities could be carried out as between one commodity taken as standard and each of all exchangeable commodities taken in turn. We should thus obtain the prices of all commodities in terms of the standard. If such a commodity were readily acceptable in exchange for all commodities, people would be willing to take it, not only because they wanted to consume it, but also because they wanted to use it to obtain other commodities in exchange. The latter might well become the principal reason why people wanted to acquire it, and there seems no reason why we should not work the apparatus using a commodity which is wanted only because it is readily acceptable in exchange for all other commodities. Such a commodity has then become money.

If we adopt money as the standard commodity in terms of which all prices are to be measured, we can use the more reasonable idea of a money income and measure both market and personal valuations, i.e. prices and marginal rates of substitution in terms of money. We need no longer worry about the possibility that the marginal rate of substitution of any one commodity for money may not diminish with an increasing proportion of total income devoted to acquiring the commodity. If the M.R.S. should not diminish, the indifference curves will not be concave upwards, and if they do not have this shape, there is no possibility of a resting-point on any exchange path across the indifference map.

If the indifference curves were concave downwards and an individual found himself at the point where a price line was tangential to one of them, it would be advantageous to move in either direction rather than to stay at that point. The individual in these circumstances could find no resting-point except on one of the axes, so he would either buy none of the commodity or would spend all his income on it, and it would appear to be largely a matter of chance which he happened to do.

If the indifference curves were straight lines, again there would be no stopping-point except on one of the axes, and which

axis was selected would depend on whether the price line cut the indifference line from below to above, in which case the stopping-point would be on the horizontal axis, or from above to below, in which case it would be on the vertical axis. There is also the possible case in which the price line and a straight indifference line are parallel, when there will be no reason why

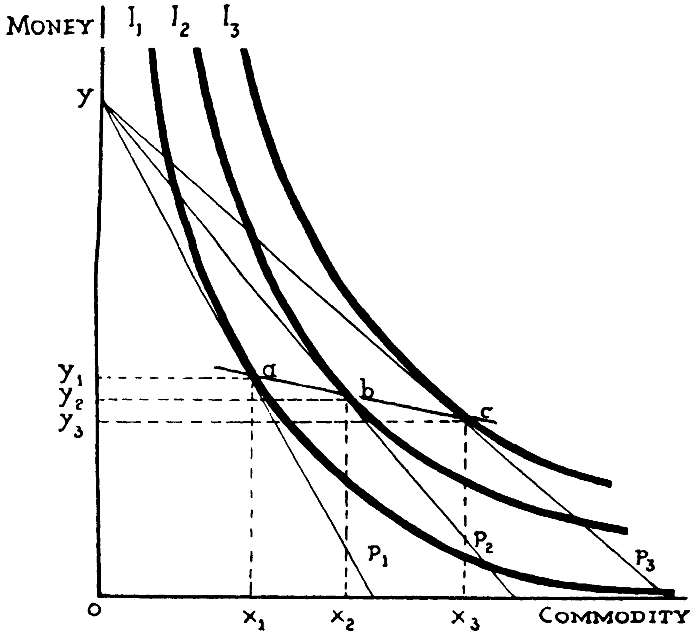


FIG. 8A.

the individual should move from any point in which he happened to find himself, since movement along the price line would leave him no better off and no worse off.

The Demand Schedule for an Individual

To determine the quantities a given individual would purchase at any one of a series of prices, let us consider that the three curves, I_1, I_2, I_3 in Fig. 8 are part of his indifference map. He has a money income of y per annum. The question then is what will be the quantities of this commodity the individual will purchase at a series of prices p_1, p_2, p_3? These

prices are represented by the slopes of the lines drawn through the point y, which touch the indifference curves at a, b and c respectively. As we have already seen, these points of tangency represent the limits to which exchange will go, and the co-ordinates of each of them represent the quantity of commodity acquired, x_1, x_2, x_3, and the quantity of income remaining y_1,

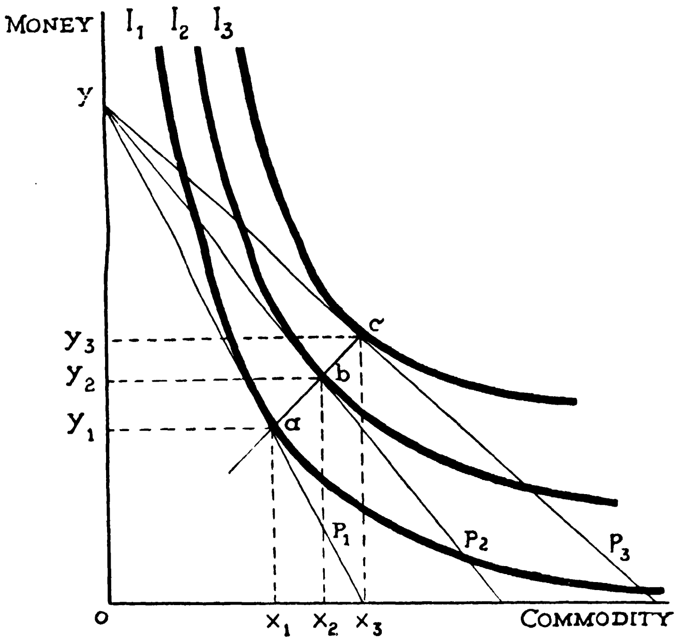


FIG. 8B.

y_2, y_3. For instance, at price p_1 the individual will buy x_1 of the commodity per annum, spending $y - y_1$ on it and having y_1 left to spend on other things.

In each of the two diagrams of Fig. 8, p_3 represents a lower price than p_2, and p_2 than p_1. In both diagrams the quantity of commodity acquired increases as price falls, but while in Fig. 8A the amount of income spent on the purchase of this commodity also increases for a price fall, in Fig. 8B, although more commodity is purchased, less is spent upon it, and the quantity of income remaining, when purchase is complete at price p_2, is

greater than what is left after the purchase has been made at price p_1.

Thus in Fig. 8A at price p_1 a quantity of commodity x_1 is purchased for an expenditure $y - y_1$. In this case an increasing quantity is purchased for an increasing expenditure. In Fig. 8B we again have x_1 purchased at p_1 for an expenditure of $y - y_1$ and a greater quantity x_2 purchased at p_2 for $y - y_2$. But here $y - y_2$ is less than $y - y_1$, and expenditure falls as price falls and the quantity purchased increases.

The actual result in any particular case depends on the shape of the indifference curves, that is, on the scale of diminishing marginal rate of substitution, and it is clearly possible that a third case could arise where expenditure would remain constant irrespective of price.

Each point of tangency between an indifference curve and a price line gives us one item in the demand schedule of the individual. At price p_1, quantity x_1 is purchased; at price p_2 quantity x_2 and so on. As we have already seen, the demand schedule can be derived directly from experience, but in this construction of a schedule on the basis of the theory of choice, we have been able to observe the separate effect of the preferences of the consumer, his income and market price, in a manner impossible by direct observation of schedules.

The Effect of Income Changes

Having examined what happens when a consumer's preferences are given, income is taken to be constant and price alone is allowed to vary, we must now assume price to remain constant and allow income to vary. In Fig. 9 let M_1, M_2, M_3 represent three different values of money income. Through each of these points is drawn a line whose slope represents the market price p, so that the three lines will be parallel. Each of these straight lines is tangential to one indifference curve at points d, e, f respectively. In Fig. 9A the points of tangency are such that point e, where M_2 is tangential to I_2, is to the right of d, where M_1 is tangential to I_1 and similarly f is to the right of e. In this case the consumer increases purchases of the commodity in question as income grows, but in Fig. 9B e is to the left of d, and f to the left of e, and the quantity of commodity purchased decreases as income increases.

Of these two cases the first is the general case and the second the special case of an inferior good. The indifference curves are those of a cheaper inferior substitute for a superior article, and people are willing to spend some of an increase in income on securing a supply of superior quality rather than one of greater quantity. We have already mentioned the case of the

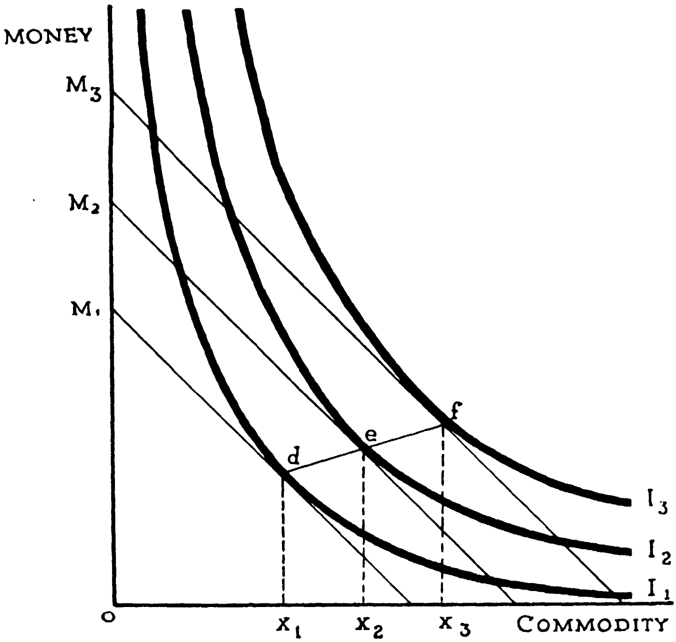


FIG. 9A.

cheapest common foodstuff in any community. People use an increased income to provide a more delicate as distinct from a more ample diet. So an increase in income induces people to buy first-class travel tickets instead of third rather than to travel more frequently, buy better seats at the theatre instead of going more frequently in the gallery.

With commodities other than inferior goods, the extent to which consumption is expanded when income increases depends on the rate of fall in the importance a consumer attaches to successive uses for an increment of commodity. If the consumer has an unsatisfied use for a further unit of one com-

modity, which is but little less important than the last one to be satisfied, an increase in income is likely to lead to increased consumption. Where there is a considerable drop in import-

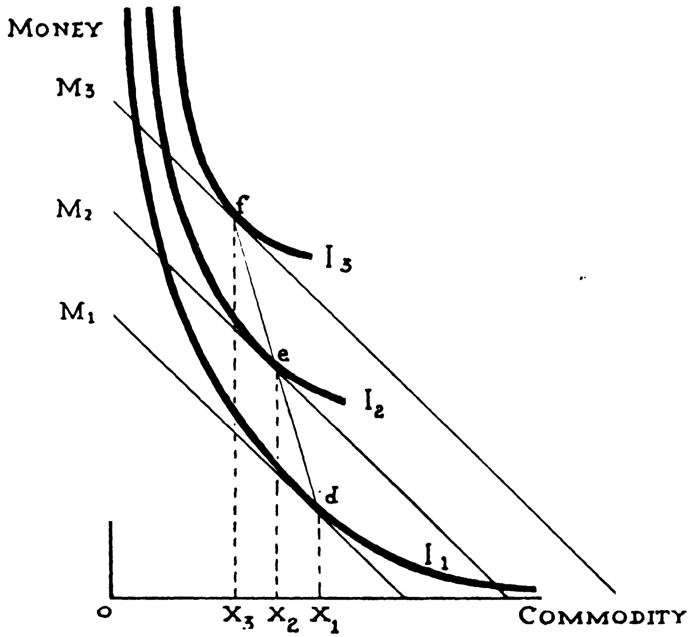


FIG. 9B.

ance between the last use to be satisfied and the first to go unsatisfied, a considerable increase in income may take place before consumption increases at all.

Market Demand Schedules

Having arrived at the demand schedule of a single individual for each commodity he consumes, we can imagine the process repeated for all individuals in the market. We can add together the quantities which all individuals will take of a particular commodity at each of a given set of prices and so obtain a market demand schedule. In order to do this, it is not necessary that the indifference maps for all of them should be alike. The result would be perfectly valid, if our market consisted of a group of persons for whom the commodity was

an inferior good and a group for whom it was not. The resulting demand schedule might, however, present some puzzling features, and in any practical problem to which such a schedule was relevant it would probably be necessary to try to isolate the two groups and to study them separately.

More Advanced Reading

Hicks, J. R.: *Value and Capital*, Chaps. 1 and 2.
Pareto, V.: *Manuel d'économie politique*, Chap. 4.

Chapter VI

PROPERTIES OF THE DEMAND CURVE

Elasticity of Demand

THERE is still more information to be gained from an examination of small changes in price and quantity demanded; as yet we have no measure of the relative sizes of change in price and change in quantity for various commodities. If price changes, quantity purchased almost invariably changes in the opposite direction, but the amount of the change will not be the same in all cases. It is clear, of course, that we cannot compare absolute price and quantity changes because of difference in units. It would be ridiculous, for instance, to make a comparison between the effects of a penny per pint change in the price of beer and a penny per ton change in the price of coal. If we use percentage changes, however, this difficulty is avoided.

A decrease in the price of a reel of cotton from $2\frac{1}{2}d.$ to $2d.$ is not likely to induce anyone to buy very much more cotton. The uses to which cotton is already being put are very important, but there are few slightly less important uses to which cotton thread might be put at the lower price. Similarly, a 20 per cent. fall in the price of table salt would hardly induce us to salt our food more heavily than we do at present. On the other hand, a fall in the price of strawberries from 1*s*. 3*d*. to 1*s*. per pound might be expected to lead to a considerable increase in consumption. Elasticity of demand is a measure of the responsiveness of quantity demanded to a change in price, and it is defined as the ratio:

$$E = \frac{\textit{Proportionate change in quantity demanded}}{\textit{Proportionate change in price causing change in quantity}}.$$

Thus, if a change in price of 1 per cent. is accompanied by an equi-proportionate change in demand in the opposite direction, both numerator and denominator of this fraction will be equal

to one, and we say we are dealing *with a case of unit elasticity*. But the response in quantity demanded to a change in price need not be equi-proportionate. Suppose that it is less than proportionate, so that a change in price of, say, 2 per cent. is accompanied by a change in demand of 1 per cent., elasticity of demand is then equal to a half. So long as the value of elasticity is less than one, demand is said to be inelastic, that is to say, it is relatively unresponsive to a price change. On the other hand, if the proportionate change in quantity is greater than the proportionate change in price, elasticity is greater than unity, and demand is said to be elastic and it is relatively responsive to price changes.

The careful reader will have noted an inaccuracy in this definition, which we must now proceed to correct. The changes in price and quantity are, in general, of different signs, an increase in one corresponds to a decrease in the other, so that elasticity of demand as defined above will always have a negative value. In order that it shall be a positive quantity it is customary to multiply the ratio by -1.

But, even so, the definition of elasticity stated above is still only a first approximation, and indeed covers two distinct measurements of the elasticity of a demand curve. The one, known as arc elasticity, deals with finite price and quantity changes, while the other, known as point elasticity, deals with infinitesimal changes. The former is so little used that we shall dismiss it with this brief mention and confine our attention to point elasticity. Then, if quantity q is being sold at price p, a price change of Δp will be accompanied by a change in quantity of Δq in the opposite direction, so that Δp and Δq will always be opposite in sign. We can then write elasticity of demand, E, as:

$$E = -\frac{\frac{\Delta q}{q}}{\frac{\Delta p}{p}}.$$

To make our original verbal definition correspond numerically to this, it is only necessary to amend each line to read, "A very small proportionate change."

We may now enquire what characteristics of a commodity

determine the elasticity of demand for it. Elasticity depends on one characteristic only, the ease with which substitutes can be found for the commodity, or with which it can be substituted for other commodities. Commodities and services are demanded because they gratify particular appetites, and in general appetites can be gratified in alternative ways. Thus the cinema, the public-house, the football match and the dog-racing track offer alternative ways of finding a brief escape from the drabness of everyday life for a large part of the population. Expenditure on house rent may be alternative to expenditure on recreations outside the home. Even hunger can be satisfied in a variety of ways. If the price of char-à-banc tours is raised, people may prefer to turn over expenditure to the theatre, or they may spend it on tennis-club subscriptions. The alternative may differ widely from the thing for which it is substituted.

We should expect to find, as is indeed the case, that the demand for luxury goods is in general elastic, because the desires they satisfy are generalised and can be gratified readily in alternative ways. Necessities, on the other hand, are goods or services which satisfy very particularised desires; substitution is difficult and may be impossible; demand is therefore inelastic. In the case of water we should in the last resort be willing to spend the whole of our income on the minimum quantity of water necessary to support life, if only on the ground that it is less painful to die of hunger than of thirst. Under more normal conditions there is a limit to the amount a given household will wish to consume for all purposes, and drastic reductions in the water rate would probably be necessary before it was considered worth-while to find new ways of consuming water. Demand for the cheapest common foodstuff is always very inelastic. Because it is the cheapest, no substitute is available if its price rises. Indeed, it is possible for the circumstances to be such that the demand curve curls round to the right as prices rise. Suppose we have a community living mainly on potatoes, which are the cheapest form of food. If expenditure on potatoes forms a large part of total expenditure and the price of potatoes rises, it may be necessary to curtail expenditure on choicer and more expensive foodstuffs and to spend more on potatoes in order to satisfy hunger. In these

circumstances, a part at least of the demand curve could rise from left to right, even though the prices of all other commodities remained the same.

The question of availability of substitutes is partially a technical question and partially a matter of price. Silver is technically a better material for the manufacture of electrical conductors than copper, but its technical superiority is not great enough to make it a substitute for the purpose at normal prices. If the price of silver fell low enough relatively to the price of copper, we might expect to see cable-makers changing their raw material.

Consumers' Outlay

We have already noted that expenditure on an individual commodity may increase, decrease or remain constant as price falls and the quantity of commodity purchased increases. Consumers' expenditure or outlay is the same thing as gross receipts from the sellers' point of view, so that this relationship is important, both in relation to the problem of distribution of consumers' expenditure and to that of the determination of the gross incomes of groups of producers. We must therefore investigate farther the relationship between changes in price and changes in expenditure. Suppose that a quantity q of a commodity is being sold at price p. If we take equi-proportional changes so that price falls by 1 per cent. and demand extends by 1 per cent., consumers' outlay after the price change is:

$$\frac{99}{100}p \times \frac{101}{100}q = \frac{9999}{10000}pq = pq \text{ approximately.}$$

Thus, when demand is of unit elasticity, consumers' outlay is the same whatever the price.

Suppose a decrease in price of 1 per cent. is accompanied by a rise in quantity of 2 per cent., consumers' outlay, which was pq before the change, is now:

$$\frac{99}{100}p \times \frac{102}{100}q = \frac{10098}{10000}pq, \text{ or } \frac{101}{100}pq \text{ approximately.}$$

Conversely, if price had risen and quantity had fallen, consumers' outlay would have decreased to $\frac{99}{100}pq$ approximately.

Using our first approximate formula for elasticity of demand, we have in this case $E = 2$.

Thus, it would appear that with elastic demand consumers' outlay increases for a fall in price and decreases for a rise in price.

Finally, let price fall by 2 per cent. and demand extend 1 per cent.

Consumers' outlay after the price change is:

$$\frac{98}{100}p \times \frac{101}{100}q = \frac{9898}{10000}pq = \frac{99}{100}pq \text{ approximately.}$$

Similarly for a price rise, consumers' outlay increases to $\frac{101}{100}pq$. Thus, it would appear that when demand is inelastic, consumers' outlay rises when price rises and falls when price falls.

Arithmetical examples do not, however, prove propositions, and those who desire rigid proofs are referred to the appendix at the end of this chapter.

The information we have extracted from these examples forms the foundations of a most useful diagrammatic analysis and enables us to draw curves of known elasticity. First, consumers' outlay on a quantity Oq at price Op is represented by any such rectangle as $OpAq$, Fig. 10, formed by the axes and the co-ordinates of the point A on the curve. Using the knowledge that, in cases of unit elasticity, consumers' outlay is constant, we can draw curves which show unit elasticity throughout their length. A series of rectangles of equal area $OpAq$, Op_1Bq_1, Op_2Cq_2 etc. (Fig. 11A), is drawn, and the points A, B, C, etc., are joined to form a curve which is a rectangular hyperbola and represents a demand curve, the elasticity at every point of which is equal to unity. Similarly, by drawing

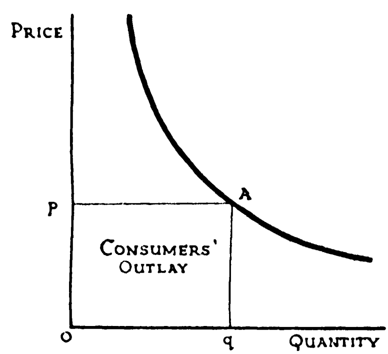


Fig. 10.

a series of rectangles $OpLq$, Op_1Mq_1, Op_2Nq_2 such that each rectangle is of lesser height and greater area than the one before it, so that outlay is greater at a lower price than at a higher price, we obtain a curve at every point of which demand is elastic (Fig. 11B). Again, the series $OpRq$,

ELASTICITY OF DEMAND AND EXPENDITURE

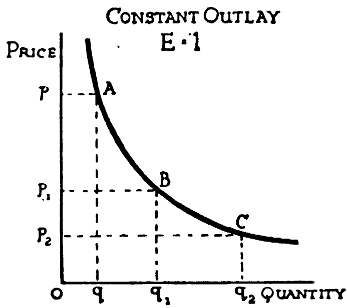


FIG. 11A.

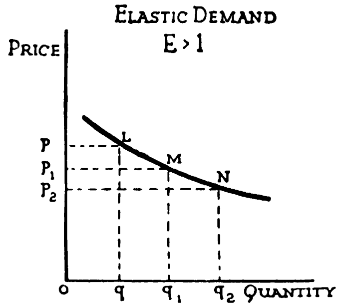


FIG. 11B.

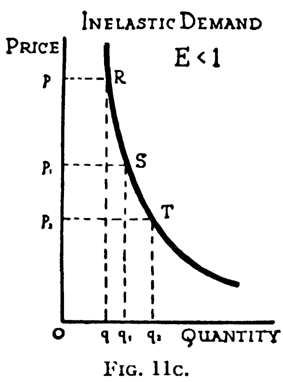


FIG. 11C.

Op_1Sq_1, Op_2Tq_2, in which area decreases as height decreases, gives a curve at every point of which demand is inelastic (Fig. 11C).

Because in diagrams drawn in this way an inelastic demand curve is more steeply inclined than the curve with greater elasticity, it is often concluded that elasticity can be judged from slope. It is true that if two curves are in close proximity the steeper curve has the smaller elasticity, but it must be remembered that the slope of the curve depends on the absolute

changes in price and quantity, while elasticity is dependent upon proportionate changes. It is, however, always possible to decide whether a given curve shows elastic or inelastic demand by constructing a constant outlay curve to cut it in the section of the curve it is desired to test.

In Fig. 12A a constant outlay curve *CO* and another demand curve *DD* cut at point *a*, and it is possible to show that curve *DD* shows elastic demand in the neighbourhood of the point of intersection. At price *op*, quantity *od* is sold and outlay is

ELASTICITY OF DEMAND AND EXPENDITURE

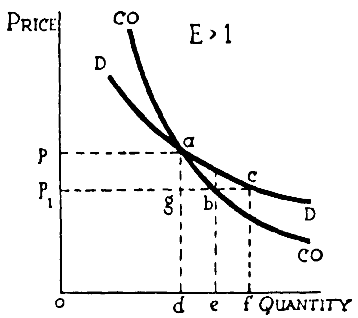


FIG. 12A.

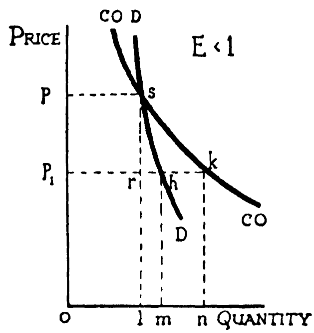


FIG. 12B.

opad. At price op_1 under constant outlay conditions, quantity *oe* would be sold and outlay would be op_1be. Under conditions of demand shown by the curve *DD*, the quantity sold at op_1 would be *of*, and the outlay op_1cf is greater than op_1be and therefore greater than *opad*. In the conditions of demand shown by *DD*, outlay increases as price falls from *op* to op_1, which from our arithmetical examples indicates that demand is elastic.

This fact can also be shown from the diagram. Elasticity on the *CO* curve at point *a* is $\frac{gb}{od} \cdot \frac{ad}{ga}$, and we know that this is equal to unity. Elasticity on *DD* at the same point is $\frac{gc}{od} \cdot \frac{ad}{ga}$, and since *gc* is greater than *gb*, elasticity on *DD* at point *a* must be greater than unity, that is, demand in the region of *a* must be elastic.

A similar argument shows that DD in Fig. 12B shows inelastic demand, and that outlay falls for a price fall. Again, $opsl$ and op_1kn are equal in area, as they are drawn to the same constant outlay curve. op_1hm is less than op_1kn and therefore less than $opsl$. Therefore, on the demand curve DD a fall in price leads to a decrease in outlay. Again as before, elasticity of demand on the CO curve is $\frac{rk}{ol}.\frac{ls}{rs}$ and elasticity of the curve DD is equal to $\frac{rh}{ol}.\frac{ls}{rs}$. The change in quantity rh is less than the change in quantity on the CO curve, so that the value of elasticity is less than 1 and DD shows inelastic demand in the neighbourhood of s.

From this we can derive a rule for judging elasticity. If a demand curve cuts a constant outlay curve from below to above, moving from left to right, demand is elastic in the neighbourhood of the point of intersection. If it cuts from above to below, moving from left to right as in Fig. 12B, demand is inelastic in the neighbourhood of the point of intersection.

Elasticity will not usually be constantly greater than unity, or constantly equal to or less than unity, throughout the length of a demand curve. It is probable that elasticity is usually greater at the upper or left-hand end than at the other end of a curve. If we think of the demand curve for a raw material like cotton or tin, demand is inelastic over most of the range of price experienced, but there is no doubt it would become elastic if price rose to levels at which substitutes which are technically superior became competitive.

Statistical Determination of Elasticity of Demand

We have already seen that the conditions on which the demand schedule is based change so frequently that the statistical determination of a schedule is difficult if not impossible. To derive such a schedule satisfactorily from direct observation, it would be necessary that the price of our commodity should vary while the prices of all other goods and services remained constant. On the other hand, if the price of our commodity moved in exactly the same way as all other prices, it is possible that consumption would remain constant. It may, however, be possible to compare proportional changes in consumption

and proportional changes in the price of the commodity in which we are interested when that price is measured in relation to the prices of goods in general or to the prices of related commodities. By such a study one might arrive at a series of computations of elasticity of demand for the commodity when its price moves in relation to other prices.

Computations of elasticity of demand have been made for a number of commodities, and though no high degree of precision can be claimed for the figures obtained, the results do at least give us some assurance of the reality of the quantity and an indication of its magnitude in particular cases.

Appendix

The demand curve for a commodity can be represented by the function:

$q = F(p)$, where q and p are inversely related.

The elasticity of demand is equal to:

$$E = -\frac{\frac{\Delta q}{q}}{\frac{\Delta p}{p}} = -\frac{p}{q}.\frac{dq}{dp}$$

Now, if $E = 1$:

$$-\frac{p}{q}.\frac{dq}{dp} = 1 \qquad \therefore p.\frac{dq}{dp} = -q$$

$$\therefore p.\frac{dq}{dp} + q = 0 \qquad \therefore \frac{d}{dp}(pq) = 0$$

$$\therefore pq = \text{constant}$$

That is, the demand curve is a rectangular hyperbola, and consumers' outlay is constant when elasticity of demand is equal to unity.

If $E > 1$:

$$\therefore -\frac{p}{q}.\frac{dq}{dp} > 1 \qquad \therefore -p.\frac{dq}{dp} > q$$

$$\therefore -p.\frac{dq}{dp} - q > 0 \qquad \therefore \frac{d}{dp}(pq) < 0$$

Where $E > 1$, consumers' outlay decreases for a rise in price

(dp + ve) and rises for a fall in price (dp − ve). Similarly, it may be proved that where $E < 1$, consumers' outlay decreases for a fall in price and increases for a rise in price.

More Advanced Reading

SCHULTZ, H.: *Theory and Measurement of Demand*, Chap. 1.

SCHULTZ, H.: *Statistical Determination of the Laws of Supply and Demand*, Chaps. 1–3.

STIGLER, G. J.: *Theory of Price*, Chaps. 5 and 6.

Chapter VII

CONDITIONS OF SUPPLY

The Reserve Price

It is already evident, from our earlier discussions, that two types of consideration determine the conditions of supply. First, we have the subjective considerations on which, so far, the chief stress has been laid. Just as the person who contemplates buying a commodity weighs the importance to him of the thing to be acquired against the importance of the collection of goods and services he will have to refrain from purchasing in order to acquire it, so the would-be seller weighs the importance to him of what he has to sell against the importance of what he can buy with the proceeds of the sale.

In one income stratum it may involve a conflict between the rival attractions of an inherited collection of old masters and a racing stable, in another it may be between reluctance to part with some household treasure and the desire to provide a child with extra food or clothing. This kind of choice exists, not merely in the unimportant cases of the supply conditions of old masters, antique furniture, period houses and other irreplaceable goods, but also in the much more important cases of the supply conditions of labour, land and capital.

The supply of any productive activity will depend on the importance to the supplier of the results of that activity. The supplier of any labour service will set against the effort demanded of him, not only the income he thereby obtains, but also the opportunities for pleasurable activity the occupation affords. The owner of land will have to choose between the income he can derive by letting it for rent and the loss of enjoyment of its amenities. If the possessor of a stock of money contemplates lending it, he has to make a choice between the possibilities of income from interest and the loss of convenience and security he incurs when he parts with control of his funds.

Secondly, we have the physical or technical aspect, the whole array of technical facts relating to the quantities of resources in existence. Technical facts concerning climate are relevant to questions concerning the supply of agricultural crops, others concerning the earth's geological structure to questions of supplies of mineral resources, while facts provided by many of the natural sciences may be relevant to the problems of the manufacture of some object of everyday use.

On these technical factors depend the quantities of productive resources required to make a given output of a particular commodity. The geological structure of a certain mineral deposit and the current state of mining technique determine the quantities of labour, machinery, fuel and stores needed to extract a ton of ore. Similarly, the expenditure of resources needed to produce a ton of an agricultural crop will depend on the natural conditions of soil and climate and on the particular technique employed.

The two sets of conditions, subjective and technical, will together determine the money costs of production of the product.

The Subjective Aspect

In Chapter IV, when we were examining the conditions which determine the quantity of a commodity a given individual would demand, we saw that the quantity demanded at a given price was the quantity which made the marginal rate of substitution for money equal to the price. This is just as true when an individual already has a stock of a commodity and wishes to retain a part and exchange a part for other goods or, what amounts to the same thing, sell part of the stock for money. The supply such an individual is willing to put on the market is equal to the difference between his total stocks and the quantity he wishes to retain for his own use, and the market supply is the sum of individual contributions determined in this way.

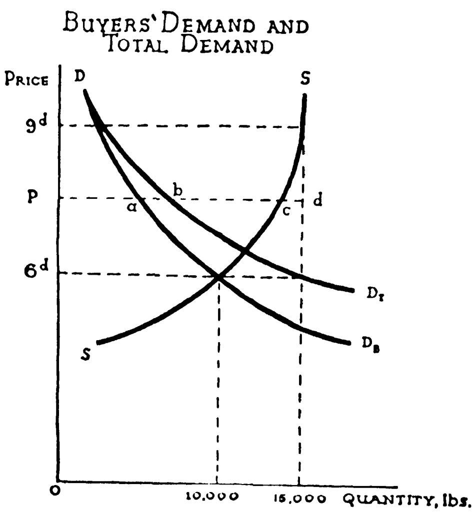
At certain prices the marginal rate of substitution of the holders may always fall short of price, however small the quantity retained. In this case the whole supply would be sold. Lower rates for the commodity might fall within the scale for marginal rates of substitution, and part of the stock would then be retained and the remainder sold.

To illustrate the matter let us take an example from the time-honoured fish market, where the catch is 15,000 lb. The sellers have a demand for their own product—to store, cure or to eat themselves, it does not matter which.

TABLE 2

Price per lb.	*Buyer's Demand* lb.	*Seller's Demand to Hold* lb.	*Total Demand* lb.	*Supply* lb.
9*d.*	7,000	nil	7,000	15,000
8*d.*	8,000	1,000	9,000	14,000
7*d.*	9,000	3,000	12,000	12,000
6*d.*	10,000	5,000	15,000	10,000
5*d.*	14,000	6,000	20,000	9,000
4*d.*	15,000	7,000	22,000	8,000

If the whole quantity in existence had to be sold without reserve price, equilibrium would be established at 4*d.*, the price at which buyer's demand is sufficient to clear the market.



FIG

Since the sellers also have a demand for their product, equilibrium is established at the point at which total demand, i.e. buyer's demand plus seller's demand, clears the market. This occurs at price 6*d.*, where the buyer's demand of 10,000 lb. and a seller's demand of 5,000 lb. together account for the whole of the 15,000 lb. in existence. We assume that, at 9*d.* per pound, the sellers prefer to dispose of the whole of the catch, but that, as price falls below this level, the quantity they prefer not to sell increases. Expressing the same data in diagrammatic form

(Fig. 13), DD_B represents buyer's demand, that is, it is the same curve we have already become accustomed to using. DD_T represents column 4 in the market schedule, so that seller's demand is represented by the horizontal distance between the two curves. Price is determined at the point where the total demand curve cuts the vertical line representing the fixed quantity in existence, and at this level 10,000 lb. are purchased. SS is the ordinary supply curve and is vertical above the price of 9*d*. Horizontal distances between it and the vertical line representing the fixed supply are equal to the demands to hold at the respective prices. Thus, at price *op*, demand to hold is represented both by the distances *ab* and *cd*. Market price is determined at 6*d*. Buyer's demand plus seller's demand equals 15,000 lb., equals supply plus seller's demand.

From this analysis it appears that conditions of supply are not entirely independent of conditions of demand, as those nineteenth-century economists who first used the demand and supply curve apparatus taught. Wicksteed, to whom this analysis is due, denied that there is such a thing as a supply curve, and insisted that it is just the obverse view of the seller's demand. This concept of market supply as the surplus which owners do not wish to retain at a given price presents no difficulties when we are dealing with a fixed stock of an irreplaceable commodity, but its relevance to a commodity which is being produced currently is more difficult to see.

If, however, we keep the idea of substitution firmly fixed in our minds, the matter is plain. Every material, every unit of labour, land or capital employed, has alternative uses, which are ruled out once the decision to devote it to one particular purpose is taken. Not only that, the owner of any productive resource has the choice between using his property to earn a money income and of refraining from so using it.

The owner of the factor, labour, can use his time and effort for remunerative work or for his own recreation; the landowner can rent out his land or use it for his own pleasure; the owner of liquid capital can keep it as cash in a strong box or can lend it at interest.

Since units of factor have alternative uses, what can be earned in one line of production cannot be unrelated to what can be earned elsewhere. Let us imagine a community which

produces only wheat and coal and which uses the same tools for either purpose. We will suppose that a day spent coal-digging and a day wheat-growing are equally arduous. Then a coal-miner will not give for his wheat supply more than the amount of coal by which his output would fall if he decided to produce that amount of wheat for himself. The amount of coal a man produced in a day would thus sell for $\frac{1}{365}$ of the annual output he could have obtained if he had decided to grow wheat instead. In the more complicated circumstances of our own world, a producer will not continue to use resources to produce a commodity if those resources might be more profitably employed elsewhere.

The owners of a factor of production will require the same income from lending out a portion of their property, whatever the use to which it is put. The reserve price for each commodity will thus have to include a payment to each unit of factor equal to what it could earn in any other line of production. If any inequality of reward occurs, units of factor will tend to move from lower-paid to better-paid occupations to wipe out the divergence.

Thus taking the total quantity of factor in existence as given, the supply of it in one line of production is only the obverse view of the demand for it in other occupations. Consequently, if the demand for a given commodity increases and no unemployed units of factor exist, the producers will have to offer higher prices to attract units of the factor they need away from other occupations. The reserve price for large quantities of a commodity will tend to be higher than that for small quantities, because it has to cover the higher payments necessary to attract resources away from their previous occupations. In the case of a commodity in current production, it is possible to regard reserve price as being determined by demands to use the factors of production employed in alternative ways.

The Technical Cost Scale

When we were investigating the nature of demand, we found ourselves much concerned with the terms on which people were willing to exchange commodities one against the other, that is, with the rates at which they were willing to transform one commodity into another by exchange. It may be fruitful, there-

fore, to examine the rates at which it is possible to transform *one commodity into another by production. It is clear that* we shall have to take into account not only the rates at which productive resources can be transformed into a particular commodity, but also the rates at which the supply of one commodity can be increased by decreasing the supply of another commodity and devoting to the production of the first the resources which would have gone to make the second.

Factors of production can, to a considerable degree, be substituted one for another and, because of the existence of scarcity, the allocation of resources to the production of our commodity means that there are less resources to make something else.

The physical relationship between output of product and the quantity of factor employed we term the technical cost scale. It is the counterpart on the supply side of the preference scale on the demand side.

In considering the change in the reserve price of a commodity as the output of it is increased, we took no account of the possibility that the physical relationship between output and quantity of factor employed may not remain the same.

In our coal-wheat example we made the tacit assumption that if one miner produced 1 ton per day, fifty would produce 50 tons and a thousand 1,000 tons. This is not necessarily or perhaps generally the case. The thousand might produce more or less than a thousand times the output of one.

The question of size of output and quantities of factor of production used per unit of output involves two distinct problems.

First, we have the problem of changes in the proportions in which factors are combined in a productive process. Most productive units, whether factories or farms, are usually designed with some particular rate of output in view. We enquire what the changes in the quantities of factor consumed per unit of output are, when actual output differs from the designed output. In the case of agriculture, the problem would take the shape of examining the effects of more or less intensive cultivation of a farm of a given size; in another sphere it might be concerned with variations in output in an automobile works where cars are built individually by workshop methods. By

giving the more highly skilled turners and fitters more assistance of labour or machines, it may be possible to increase output above normal. By giving them less assistance, the rate of output may be lowered. We then discover how output per unit of each factor varies as more or less of the factor is used.

The problem may be most conveniently discussed by making the arbitrary assumption that the quantity of one factor of production employed is fixed, and that the quantities of the others are varied in turn. The law relating input of variable factor and output of product in these conditions is named, perhaps not very accurately, the Law of Diminishing Returns.

In the second problem we are concerned with changes in the whole scale of operations. Where formerly it was a question of the combination of factors used on a holding of a given size, we are now examining the difference in rate of return to factors, when holdings are of 10 acres, 500 acres, 1,000 acres and 10,000 acres, and entirely different techniques of organising factors are employed. Similarly, in the automobile example it is a question of the difference in the amounts of labour and capital required to make a car when the workshop technique is employed to produce 100 cars per year and when a mass-production method, with its assembly line fed with components brought by conveyers from specialised workshops, is turning out as many per day. It is a common error to imagine that a larger output can invariably be produced with a smaller expenditure of factors per unit than a smaller output. It is possible that, as the scale of production increases, a greater expenditure of factors per unit of output becomes necessary.

A clear line of demarcation between cases of this type and those where the Law of Diminishing Returns is operative is that changes in scale are supposed to involve such major changes in technique as substituting highly specialised machinery of an automatic type for machine tools, and a labour force of machine-minders for one of skilled craftsmen.

Where a change in the scale on which production is organised in a particular industry takes place, the process of growth will in general take time, while the different organisations of factors envisaged by the Law of Diminishing Returns are to be regarded as alternatives at a point in time.

In the present chapter we are concerned with the case where the different organisations of factors are considered as alternatives at a moment of time.

Diminishing Returns and Variable Costs

The Law of Diminishing Returns is commonly stated thus: *If the quantity of a factor of production, used in co-operation with a fixed quantity of another factor, is continually increased, output per unit of the variable factor increases, reaches a maximum and then declines.*

Instead of looking at it from the point of view of output of product per unit of factor input, we can state it in terms of quantity of factor required per unit of output. The statement then runs that in the same circumstances input of factor per unit of output diminishes, reaches a minimum and then increases. It is a purely physical law derived from observation, being first stated for agriculture and long believed to have some special appropriateness to that industry. In the traditional agricultural example, when the labour force is very small, cultivation will be of a very cursory nature. As the labour force is increased, the improved cultivation will be rewarded with a larger crop, but if cultivation is further increased, the soil can be made too loose and the crop impaired.

The same effect is to be observed if in a factory of given size and equipment we increase the number of workers from zero. A certain minimum number will have to be employed to provide essential services before machinery can start moving. Then, as further workers are added and the process of manufacture can begin, output per head of all employed will increase, as those who provide motive power, heat, light, ventilation and ancillary services become a smaller proportion of the total. Finally, if numbers are continually increased, conveyers will not provide material or remove finished work quickly enough to keep all working at full capacity. The quantity of tools and appliances will not be enough to go round, and workers will be sharing them and wasting time waiting. Output per head will decrease. Again we have some quantity of the variable factor labour which will produce a larger output per head than any larger or smaller quantity.

The results obtained by increasing the amount of labour employed on a given piece of land are represented in Table 3. Output per head increases from 25 units for 4 men to $28\frac{1}{2}$ for 8 men, and thereafter declines. Total output increases, but not at a uniform rate, and reaches a maximum for 12 men.

TABLE 3

Labour Force	*Average Output*	*Total Output*
4	25	100
5	26	130
6	27	162
7	28	196
8	$28\frac{1}{2}$	228
9	28	252
10	27	270
11	26	286
12	24	288
13	22	286

We assume that the labourers employed are exactly alike as regards the quantity and quality of their labour, and that the order in which they are selected for addition to the labour force makes no difference to the result.

Graphical Representation of Diminishing Returns

Fig. 14 illustrates the shape of the curve showing the relationship between input of factor and total output, according to the Law of Diminishing Returns.

Let *PP* by its height show the total output when varying quantities of our factor are used in combination with fixed quantities of other factors. The curve begins on the factor axis at a distance from the origin, for we will suppose that some minimum quantity of factor must be used before any output is obtained. When *oc* units of factor are employed, output is *bc*. Average output per unit of factor is $\frac{bc}{oc}$ or the slope of the line *ob*. We can therefore represent output per unit of factor by the slope of the line joining the origin to the point on the curve vertically above the point denoting the quantity of factor in question.

Let us imagine a point travelling along the curve from ***a***, but

connected all the while with an elastic thread to *o*. As the point moves up the curve, the line joining it to *o* moves in an anti-clockwise direction, so that its inclination from the horizontal becomes greater, until the position *od* is reached. After this it rotates in a clockwise direction and its slope decreases.

Now the slope of the straight line represents the average output per unit of factor, so that this quantity is a maximum for *oe* units, *e* having the position on the horizontal axis immediately below *d*, since *od* is the position of maximum slope of a line joining the curve to the origin.

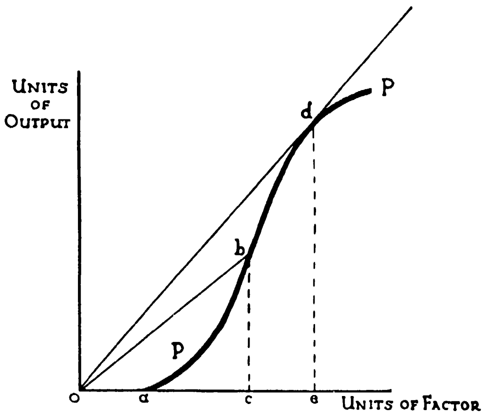


FIG. 14.

It is clear, then, that the curve *PP* reflects for total output the condition stated by the Law of Diminishing Returns, that average output increases, reaches a maximum, and then declines as input of factor is increased.

This information about the output of product and input of factor can be of immediate service to us in our investigation of conditions of supply. Let us think of an individual firm and endeavour to discover the relationship between size of output and cost per unit. We will consider the firm to be an organisation capable of producing outputs of different sizes, but being designed with some particular output in view. It will have its limitations; as output increases, it eventually finds itself working to capacity and can produce no more. It will probably possess a certain amount of land, buildings, plant and equipment; it will have a managerial organisation of a certain degree of ability, and these factors will impose limitations on the size of output which is possible, for it will not be able to increase or decrease them readily or over short periods of time.

Within the limitations imposed by this fixed organisation, it will be able to decide on a particular size of output, and buy or hire the materials, stores and labour it requires to produce that output.

We can look at the problem therefore from the point of view of varying quantities of factors applied to a fixed quantity of certain other factors and the Law of Diminishing Returns is applicable. This gives us the relationship between quantity of output in relation to input of factor, and we must turn it upside down to put it in the form of the quantity of factor consumed in relation to the size of output. That is, we want to use output of product as the independent variable instead of input of factor, as in the case of the Law of Diminishing Returns.

To do this, we must first replot Table 3, this time measuring quantities of output along the horizontal axis and quantities of labour along the vertical axis (Fig. 15A). The curve will usually

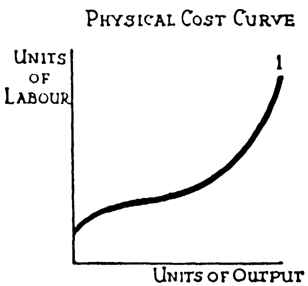


FIG. 15A.

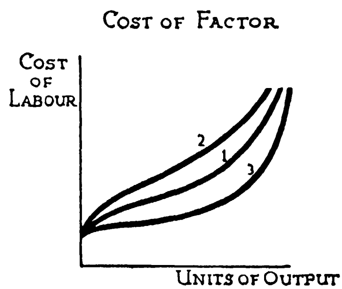


FIG. 15B.

start some distance up the vertical axis as some expenditure of factor is often required before any output is obtained. The curve rises at a decreasing rate for some time, then the rate of rise begins to increase and finally the curve becomes a vertical straight line.

This curve represents cost in terms of actual physical cost, and we can turn it into money cost by multiplying by the cost per unit of labour. If we can assume that the money cost per unit of labour remains the same irrespective of quantity purchased, curve 1 in Fig. 15B will be an exact replica of the curve in Fig. 15A. If the price to be paid for labour increases as more

is bought, curve 2 will rise more rapidly than the curve in Fig. 15A, and if the price falls with the increased quantity purchased, curve 3 in Fig. 15B will rise less steeply than the curve in Fig. 15A. In any case, the curve will still rise from left to right, for a larger quantity of labour will always cost more than a smaller quantity.

More than One Variable Factor

So far we have considered the total cost of one factor only, but our argument will hold for every variable factor. For each of them we can get a curve of the shape of Fig. 15.

If the employment of the factor is dependent entirely on the size of the output and falls to zero when the output is zero, then the total-cost-of-factor curve will start at zero. If, on the other hand, some quantity of the factor must be employed so long as the production organisation is kept in existence, even if output is nil, then the cost-of-factor curve will start at a point on the vertical axis, whose height from the origin will indicate the minimum quantity which must be used even if output ceases. In most cases a firm will be obliged to retain a certain nucleus staff in order that it can remain in existence. In deep mining, if there is any prospect of the mine being worked again it may be necessary to continue to pump and ventilate to prevent it becoming unworkable.

Addition of Factor Costs to give Total Cost

Each of these curves relates to a particular organisation of production: to a firm which employs fixed quantities of certain factors of production, such as land and buildings, plant, managerial and supervisory labour, and variable quantities of other factors, such as fuel, stores, materials, power and operative labour. We can therefore add together all the curves for the variable factors and get a curve of similar shape which will show the total cost of variable factors for any size of output. The factor costs, which are fixed irrespective of output, can also be added together, and this total, added to the cost of variable factor for each size of output, will push the curve of variable costs upward, parallel to the vertical axis.

Thus in Fig. 16, curve I represents total variable costs, and curve II total costs, so that the total variable cost of output

oa is *ab*, but the total cost is *ac*. The height of the point at which curve I cuts the cost axis is the cost of the minimum quantity of variable factors which must be used. It can conceivably be zero, in which case *d* and *o* coincide. The vertical height *de* between curves I and II represents the cost of the fixed factors. This will include such expenses as rent, interest on loans raised on fixed interest terms, and salaries fixed on long-term contracts, which have to be met whether any output is being produced or not. But there are other factors, which co-operate in the firm, but which are not entitled to any fixed reward. Nevertheless, if they are to continue to be employed in the firm in question, they must be paid as much as they can earn elsewhere, and an allowance for them must be included in fixed cost.

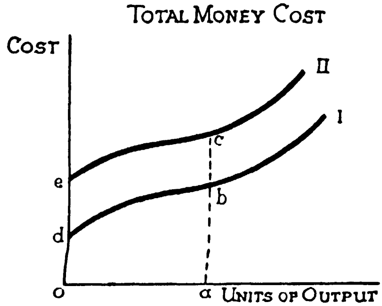


FIG. 16.

The relative importance of fixed and variable costs differs greatly from one enterprise to another.

A case where fixed charges were relatively unimportant used to exist in the woollen industry. It was at one time possible for anyone with the necessary technical ability to secure a contract to make cloth, hire the necessary floor space and machinery in a mill, engage labour, borrow the money needed for wages and yarn, and produce cloth. Even here we must regard the individual himself as a fixed factor of production, and he would probably have to have a certain amount of capital of his own before he could undertake such an enterprise.

In railway operation, on the other hand, a considerable staff must be working before a single train can run, so that *od* (Fig. 16) is relatively large. Also considerable capital is invested in the track, rolling-stock, buildings and equipment, so that *de* is also large. As the cost of the variable factors is a relatively small part of the total cost, the cost curve will rise gradually.

Average Variable Costs of a Firm

We have seen, from the Law of Diminishing Returns, that as more units of factor are applied, output per unit of factor increases, reaches a maximum and then declines. If, therefore, we invert the quantity "output per unit of factor" and so obtain "consumption of factor per unit of output," this latter quantity decreases, reaches a minimum and then increases as output is increased.

Thus the average physical cost per unit of output, say the number of man-hours per unit of product, decreases at first as output is increased, then reaches a minimum and finally increases.

Let us now suppose that labour may be bought at a constant price per man-hour, then we can say of the money cost of labour that, as output increases, the money cost of labour per unit of output decreases to a minimum and then increases. The Law of Diminishing Returns applies to all variable factors, so that we can make the same statement of all the variable factors put together so long as there is at least one factor in fixed supply, and this we have seen will normally be the case for an individual productive unit or firm.

As the output of a firm increases, the average cost of the variable factors per unit of output decreases, reaches a minimum and then increases.

Marginal Cost of Production for a Firm

Now if the average cost of variable factors per unit of output is falling—that is, if the total cost of the first two units divided by two is less than the cost of the first unit, and the cost of three units divided by three is less than half the cost of two units—it follows that the extra cost of the second unit must be less than the first and the extra cost of the third less than the extra cost of the second.

At first, then, the extra cost of a further unit will always be less than the extra cost of the one before it, but this state of affairs may not continue indefinitely.

Suppose the extra cost of a further unit begins to increase, what happens then? So long as this extra cost is less than the average of the previous units, the addition of the further unit

of output will reduce the average cost, but once the extra cost of a unit has become equal to the average, this is no longer so. If the extra cost rises above the average, then its addition increases the average, and the average too begins to increase, though it will lag behind the higher cost of the last unit added.

At this point it is convenient to give a name to this *extra cost* of a further unit of output. We call it *Marginal Cost* and define it as follows:

The marginal cost of an output of a given size is the change in the total cost when that output is increased or decreased by one unit.

We shall need to refine that definition later, but for the present it will serve our purpose. It is clear that marginal cost is affected only by variable costs, which alone enter into the extra cost of one more unit.

An arithmetical example may clarify the relationship between Variable Cost, Average Variable Cost and Marginal Cost.

TABLE 4

Units of Output	*Total Variable Cost*	*Average Variable Cost*	*Marginal Cost*
0	0·0	0·0	0·0
1	7·5	7·5	7·5
2	13·0	6·5	5·5
3	17·5	5·8	4·5
4	20·9	5·2	3·4
5	23·8	4·8	2·9
6	26·5	4·4	2·7
7	28·9	4·1	2·4
8	31·2	3·9	2·3
9	32·8	3·6	1·6
10	34·4	3·4	1·6
11	36·7	3·3	2·3
12	40·0	3·3	3·3
13	44·8	3·4	4·8
14	51·0	3·6	6·2
15	60·0	4·0	9·0

The second column is the total cost of the variable factors used to make the number of units of output shown in the first column.

The third column is obtained by dividing the second column by the first, and the fourth column by subtracting from the corresponding item in the second column the previous item in that column.

Thus the marginal cost of the third unit is the difference between the cost of two units and the cost of three units, i.e. 17·5 — 13·0 = 4·5.

It is seen that as long as the marginal cost is falling, average costs fall, but somewhat more slowly. As soon as the marginal cost ceases to decrease as at the tenth unit, the average begins to drop very slowly indeed, and when the marginal cost becomes equal to the average cost, as it does at twelve units, the average cost then begins to rise, but again more slowly than the marginal cost.

Costs and the Supply Curve

It is now time we attempted to discover how far we can use the cost output relationship for a firm as the basis of the supply schedule. We will suppose that our producer is working for a market where there is a market price which he can take as given; that is, it is unaffected by any change in his output.

It seems reasonable to suppose that a firm will try to make its net receipts a maximum, that is, it will try to get the maximum possible difference between the payments it is essential to make in order to secure an output of a given size, and the receipts obtained by selling that output at the market price. These expenses will consist of the payments to the variable factors. From net receipts it will have to meet contractual obligations to fixed factors, but as these are a fixed sum it will, by making net receipts a maximum, either maximise the payments to the factors which work for a contingent reward or minimise the losses of the operation. The owners of the factors receiving contingent rewards are, collectively, the firm, so the policy of maximising net receipts, maximises the firm's income or alternatively minimises its losses, since if net receipts do not cover contractual fixed costs, the firm will have to find the necessary funds or go bankrupt.

Of course, if a firm were faced by a market price of 4·1 and Table 4 represented its cost scales, it could, by producing 7 units, cover total costs of 28·9 (approximately) and pay every factor of production employed as much as it could expect to earn anywhere else. This output would not, however, satisfy the condition that net receipts are to be a maximum, because a further unit would add only 2·4 to costs, but it would add 4·1

to receipts. It would indeed pay to go on increasing output until an extra unit added as much to costs as it did to receipts. This would happen approximately at 12 units. It would be worth-while adding the twelfth unit but not the thirteenth.

Similarly, if 15 units were produced, every factor would be receiving its market price, as receipts and costs would be (approximately) equal, but now the reduction of output by a unit would deduct more from costs than it would from receipts, and, pursuing this process, we should again arrive at 12 units. We have the condition therefore that net receipts are a maximum when any small variation in output results in the same sum of money being added to or subtracted from total costs and total receipts so that, in a market such as we are now considering, output will be determined at the level which makes marginal cost equal to price.

This conclusion gives us the basis of the supply schedule of the firm for which we have been looking. *Given a market price, output will be that quantity whose marginal cost is equal to the given market price* because the firm is then maximising net receipts or minimising losses. The schedule is the marginal cost scale.

In the marginal cost scale, however, we have two quantities for every value of marginal cost except the lowest, and we clearly want a schedule where only one output corresponds to each price. We can at once rule out the descending portion of the marginal cost scale. If, for instance, the market price is 2·9, it will not pay the producer to limit output at 5 units, for he can produce the sixth for only 2·7. For every unit until the twelfth its production will add more to receipts than it does to costs. Given a market price of 2·9, it will pay the producer to produce the eleventh unit, but not the twelfth. Our supply schedule will therefore be the ascending portion of the marginal cost schedule.

Finally, we must decide whether the supply schedule will be the whole of the rising portion of the marginal cost scale, and here our first tentative principle again comes in useful. Although a producer will not object to producing an output if receipts from the sale exceed his total variable costs, he will not produce an output if receipts fall short of such costs, for he would be better off if he produced nothing. We must therefore look again at the proposal to produce 11 units at a price of

2·9.[1] Total variable cost is 36·7 and total receipts 31·9. It will therefore be preferable to produce nothing. If 12 units are produced and sold at marginal cost, both costs and receipts will be 40. If 13 units are sold at the marginal cost of 4·8, costs will be 44·8 and receipts 62·4. Production will now show a profit, and this situation will also hold for all outputs above 12 units. The supply schedule will therefore be the marginal cost scale above that value of marginal cost which is equal to minimum average variable cost.

The Market Supply Schedule

We have discovered the basis of the supply schedule of an individual producer, and we can obtain the market supply schedule by totalling the individual schedules. Since each individual contribution is based on a rising scale of marginal cost, the supply schedule will show a direct relationship between price and quantity; lower prices will correspond to smaller quantities and higher prices to larger quantities.

When we were discussing the construction of market demand schedules from individual schedules, we saw that different individuals had different maximum prices above which they would not buy, so that, as price fell, the demands of additional consumers were brought into market demand. So also in the supply schedule there is no reason to suppose that the marginal cost scales of different producers will be the same and the minimum price at which any contribution at all is made to market supply will differ from one producer to another. As price rises, therefore, additional producers may be drawn into making a contribution to supply.

It must not be supposed that such producers are standing idle when they are not contributing to the supply of this particular commodity; it is our picture of a producer as a specialist, making only one product, which is wrong. The equipment possessed by many producers is capable of making a wide variety of products, and is used for one product or another according to their relative profitableness. At certain price levels only the specialist firms can operate, but as prices rise, it is worth-while for some of the non-specialists to drop

[1] I.e. continuing the previous example. The argument applies *a fortiori* if price equals marginal cost of 11 units, i.e. 2·3.

some other line of production and temporarily enter this one. We see, therefore, that a supply at a particular price is affected by the prices of other products using the same factors of production, for the price at which a given producer will enter the market depends, not merely on his cost scale for the product in question, but also on the prices he is obtaining for his other products. If our commodity cannot offer his factors of production more than they are earning in their present occupations, he will not divert them to a new occupation.

Elasticity of Supply

Just as we defined a measurement of the responsiveness of the demand to price changes, so we can construct a similar measurement of elasticity of supply. It is measured by a small proportionate change in quantity divided by the small proportionate change in price which calls it forth. Since the relation between supply and price is a direct one, not inverse as in the case of demand, changes in price and quantity will always be in the same direction, both positive or both negative, and elasticity of supply will always be a positive quantity.

Just as with elasticity of demand, we speak of (*a*) unit elasticity in the case of equi-proportionate changes, (*b*) inelastic supply if the change in price is relatively the greater, and (*c*) elastic supply if the proportionate change in quantity is greater than the proportionate change in price.

Graphical Representation of Elasticity of Supply

In each of the three diagrams in Fig. 17, *SS* is a supply curve, and we wish to get a measurement of elasticity of supply at the point *a* in each case. Let *ad* be the tangent to the supply curve at the point *a*, cutting the quantity axis at *d* in each case. Then let *c* be a point, so near to *a*, that it can be considered to lie both on the curve and on the tangent to the curve. A movement from *c* to *a* on the curve means that output is increased by the small amount *cb* when price is increased by the small amount *ab*. The quantity *cb*/*ab* is dependent solely on the slope of the tangent, and remains the same if we take *c* a finite distance from *a* on the tangent. We have defined elasticity of supply as:

$$E = \frac{\text{Proportionate change in quantity supplied}}{\text{Proportionate change in price causing change in supply}}.$$

Hence elasticity of supply at *a* is equal to:

$$\frac{cb}{oe}\Big/\frac{ab}{ae} = \frac{cb}{ab}\cdot\frac{ae}{oe} = \frac{de}{ae}\cdot\frac{ae}{oe} = \frac{de}{oe}$$

since in every case *abc* and *aed* are similar triangles.

ELASTICITY OF SUPPLY

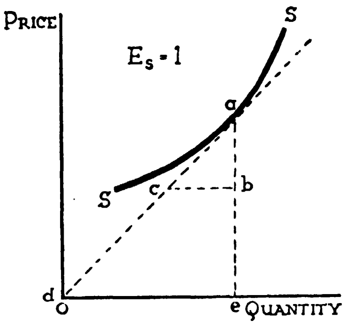


FIG. 17A.

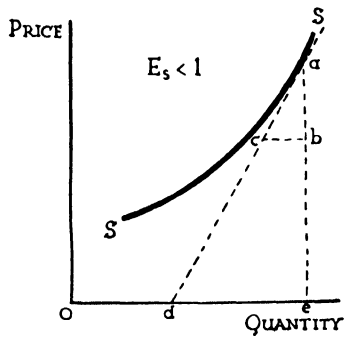


FIG. 17B.

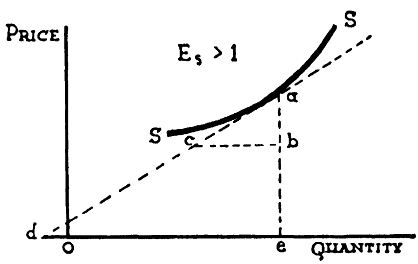


FIG. 17C.

Now in Fig. 17A, *d* and *o* coincide, so that $de = oe$.

In this case, therefore, elasticity of supply $= \frac{de}{oe} = 1$.

In Fig. 17B, *oe* is $>$ *de*, so elasticity of supply is < 1. Supply is therefore inelastic at point *a*.

In Fig. 17C, *de* is $>$ *oe*, so that elasticity of supply is > 1. Supply is therefore elastic at point *a*.

We have thus a simple test of elasticity of supply. If a tangent drawn to the curve at the point in question and produced to cut one of the axes passes through the origin, then we

have unit elasticity of supply; if, on the other hand, it cuts the horizontal axis first, supply is inelastic; if it cuts the vertical axis first, supply is elastic.

Causes of Varying Degrees of Elasticity of Supply

The basis of the supply curve is the relationship between quantity and marginal cost, so that it is to influences affecting this relationship that we must look for the causes of elasticity. First, we have the physical relationship between input of factor and output of product, governed by the Diminishing Returns Law for the particular productive process. The straighter the Diminishing Returns curve over the relevant range of output, the more constant marginal product, and hence marginal input of factor, will be. This means in concrete terms that the more easily one factor can be substituted for another in the productive process without consumption of factor rising rapidly, the greater is elasticity of supply.

But marginal cost depends, not only on the physical relationship between factor and output, but also on prices of factors. If an industry can call on supplies of factors used by other industries without bidding up the price substantially, the rise in marginal cost will be less than if large increases in payments have to be made to attract them away from their existing occupations.

The response in increased supply to an increased price depends, not only on the willingness on the part of producers, who are already contributing to market supply, to increase their output, but also on the willingness on the part of producers, who though capable of producing the commodity in question are not doing so, to make a contribution. The effect of the entry of such producers may be sufficiently important to cause the supply to become more elastic at some point in the schedule, but the general tendency attributable to the Law of Diminishing Returns will be for the supply to be most elastic at the lower end of the scale and to become more inelastic as price rises, eventually becoming perfectly inelastic when no further factors of production can be attracted away from other occupations and a rise in price no longer results in any increase in supply.

Properties of the Short-run Supply Curve

It is well to remind ourselves of the *limitations* of the supply curve we obtain from plotting the supply schedule derived in the manner described above. Like the demand curve, it describes a number of alternatives open at one particular time. Any reallocations of factors of production it entails must be capable of being effected *without an appreciable time-lag*. Furthermore, it describes the behaviour of a producer in a market in which he has no influence over the market price.

These conditions exclude a very large and important field of reality, and it is the purpose of the next four chapters to amend these rigorous conditions sufficiently to permit the examination of some of the excluded field.

FURTHER READING

HENDERSON, H. D.: *Supply and Demand*, Chaps. 4 and 5.

MORE ADVANCED READING

SCHULTZ, H.: *Statistical Determination of Laws of Supply and Demand*, Chaps. 4 and 5.
STIGLER, G. J.: *Theory of Price*, Chaps. 7 and 8.

Chapter VIII

PRICES OF FACTORS OF PRODUCTION

Demand for a Factor

So far we have been mainly concerned with demand and supply conditions for goods which are directly consumed, though we have been careful to point out from time to time that the apparatus we have been constructing is equally well suited to examine problems involving goods which are desired to make other goods. We shall now give an account of how these fit into the scheme of things, and shall include in our enquiry all those goods and services which co-operate in the production of goods and services in a consumable form, that is, those which are usually termed factors of production.

From the nature of the case, the bids of would-be possessors will depend upon the uses to which a particular factor can be put and to its technical efficiency in those uses.

The " utility economist " spoke of a factor deriving its utility from the utility of the things it helped to make. We shall think of it as capable of being transformed into other commodities, and shall regard the owners of it as virtually exchanging the use of it for a *share* of the product.

The idea of transforming a factor into a product at once involves a rate of transformation, and we have already discovered in the Law of Diminishing Returns the relationship between one rate of transformation and the quantity used. However, the importance of the marginal rate, in our discussion of supply of our commodities, leads us to look for a similar rate of productivity.

Diminishing Returns and Marginal Rate of Output

The Law of Diminishing Returns states that the average rate of productivity of a variable factor, or the average output per unit of factor, increases, reaches a maximum and then declines

as the quantity of factor concerned is continually increased, while quantities of other factors are kept constant. From this we have deduced the shape of the total product curve and the relationship between the two.

We saw in the last chapter that output per unit of factor can be represented by the slope of a chord from the origin to

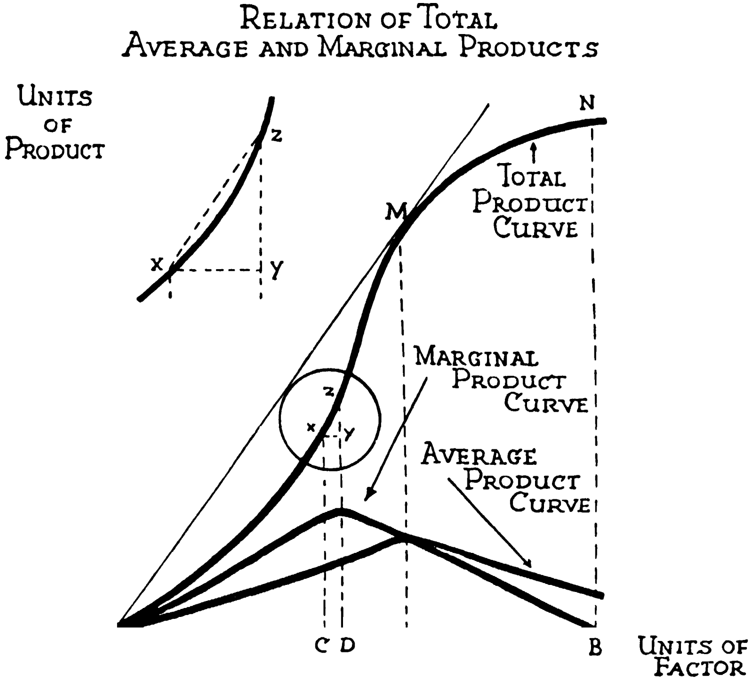


FIG. 18.

the point on the total product curve representing that quantity of factor, and that maximum average output per unit of factor is represented by that chord, which is also a tangent to the total product curve (see p. 81). So in Fig. 18 the input of factor for which average product is a maximum is OA, and the resulting total output is AM.

From this diagram we can also get the marginal rate of output. When OC units of variable factor are being used and the resultant output is Cx, an increment of factor xy results in an increment of output yz. The marginal rate of output is yz/xy,

which is also the measure of the slope of the chord between the points x and z on the curve. If xy can be taken as very small, the chord xz is identical with the arc xz and yz/xy is also the measure of the slope of the curve.

But yz/xy is the marginal rate of output when OC units of factor are employed. Consequently, the marginal rate of output for a quantity of factor can be represented by the slope of the total product curve at the point corresponding to that quantity of factor. The marginal rate of output is the ratio between a very small increment of factor and the very small increment of product it produces. Then, if we can take one unit of factor as a very small increment, we have:

The marginal product of a quantity of factor is the increase in product when that quantity is increased by one unit, or the decrease in product when that quantity is decreased by one unit.

We can now trace the relation between the marginal rate and the average rate per unit of factor.

The total product curve being concave upward at its lower end, the chord to the curve from the origin will always be sloping less steeply upward than the curve at the same point. Therefore the marginal rate, which is measured by the slope of the curve, will be greater than the average rate which is measured by the slope of the chord. Between O and M the slope of the curve increases for a time, reaches a maximum, and then decreases until at M, where the chord from O is a tangent, the slopes of the curve and the chord are equal. Therefore the slope of the curve decreases until at N it is zero, while the slope of the chord, though decreasing, is greater than the slope of the curve after the point M.

Translating this into economic quantities, we see that marginal product increases as the quantity of variable factor increases; it reaches a maximum for some quantity of factor less than OA; for OA marginal product and average product are equal, the latter being at its maximum value. As the quantity of factor increases from OA to OB, the marginal product decreases more rapidly than the average product until, when OB is being used, the addition or subtraction of another unit of factor leaves output unchanged.

The marginal product curve in Fig. 18 rises above the average curve, falls cutting the average product curve at its peak, and

continuing its descent crosses the horizontal axis at B, the input corresponding to maximum total output.

We are now in possession of the piece of information we set out to discover: the curve of marginal products or "the marginal productivity curve" is of an inverted U shape, and we can now proceed to use this knowledge in our examination of the process of factor pricing.

But before we proceed to use this new rule, let us remind ourselves that it is a purely physical law. It is derived from the Law of Diminishing Returns, and hence relates to physical quantities of factor and product.

We shall soon wish to talk of the value of the marginal product or the marginal value productivity, and we shall then have to remember that a second variable has entered into this quantity. There is not only the relationship between quantity of factor and quantity of product, but also the relationship between the quantity of product and the price per unit which can be obtained for it.

In many of the problems with which we shall have to deal, we shall be concerned with the increased use of a factor by a firm operating in conditions of pure competition. Then we can assume that any output will be sold at a constant price, and we can multiply variable marginal physical products by a constant price per unit. The marginal physical products curve will differ from the marginal value product curve only in scale. The same curve can be used for both purposes by suitably labelling the vertical axis.

If we cannot assume conditions of pure competition[1] or if we are talking of the output of an industry, the physical and value product curves are no longer identical, because a larger output can be sold only at a lower price. In the present chapter we shall assume that the output of an individual firm can always be sold at a constant price, irrespective of size of output, so that the shapes of the physical and value product curves are identical.

Let us again consider a firm which is increasing its output in moving towards equilibrium, but, instead of thinking as we did

[1] The conditions of pure competition are discussed in Chapter IX, but here it can be taken to be a state of affairs in which a given firm can sell whatever quantity of output it chooses at a constant price.

before of changes in total cost, we will think of increases in the *quantity of each factor of production used.* *We will suppose that the firm is operating in conditions of perfect competition.* The following conclusions can then be drawn.

Firstly, the firm will always be willing to hire another unit of a factor, so long as the increment of product, due to the addition of that unit, is worth more than the cost of hiring the unit. When the value of the marginal product becomes equal to the cost of the unit of factor, no more will be hired.

Secondly, the firm is hiring more than one factor, and it will not increase its use of one factor if the same increment of product can be obtained more cheaply by using more of another factor. It will therefore combine its factors in such proportions that a further pound spent on one will lead to the same increase in output as a pound spent on any of the others.

If these two conditions are satisfied, the firm is in equilibrium. It will be producing that size of output which, given (*a*) the collection of fixed factors it possesses, (*b*) the price it pays for its variable factors, and (*c*) the price it can obtain for its product, is more profitable to it than any other size of output. It will be paying for each factor a price equal to the value of the marginal product of the quantity it employs. Since a firm will always be willing to pay for a quantity of factor a price per unit equal to the value of its marginal product, we have determined the firm's demand price for any quantity of the factor.

A factor is used by many firms in the production of a variety of products, and the demands of each of them must be taken into account in determining total demands. Suppose that in a self-contained community (imports and exports being thereby ruled out), identical units of labour are used for growing wheat and sugar-beet. Then, if there is an increase in the demand for sugar-beet, the value of the marginal product of labour in sugar-beet growing will be raised because of the rise in price, and the beet-growers will raise wages to attract labour away from wheat-growing. As the numbers employed in beet-growing increase, the marginal physical productivity of labour in that occupation will fall, and as labour is lost to wheat-growing, the marginal physical productivity of labour in that occupation will rise. Furthermore, as the quantity of beet

produced increases, its price will fall, and as the quantity of wheat grown is decreased, its price will rise. For both these reasons, the value of the marginal product is falling in the case of beet-growing and rising in the case of wheat-growing. Since the marginal value product for beet was originally raised above that for wheat, they must eventually become equal once more. When values of marginal products are equal in both occupations, labour will cease to move between them and equilibrium will have been established.

Thus, *so long as units of factor are free to move, identical units in different occupations will tend to receive the same reward.*

We can therefore add together the separate demands for a factor from firms in the same or different lines of production. At any given price each firm will demand a quantity, the value of whose marginal product is equal to that price, and the total demand schedule will be the sum of the individual schedules as before.

Throughout this argument we have made the assumption that marginal productivity was decreasing, although we know from our discussion of diminishing returns that it first increases, reaches a maximum and then decreases. We can show we are concerned here only with the descending portion of the curve by an adaptation of the method we used for its counterpart, the ascending portion of the marginal cost curve. We are assuming that the quantities of all other factors remain constant, while we vary the quantity of one factor, as a producer in a competitive market will be concerned to maximise the net product of this factor, and he will employ none of the factor if there is no net product. Since we are dealing with a competitive market, where the producer can sell any quantity of product at the market price and must pay every unit of the variable factor employed the same price, it does not matter whether we speak in terms of physical product or in terms of money value.

Then, if the rectangles of Fig. 19 represent the additions made to product by successive units of factor in the case of a firm who can buy the factor at a market price *OC*, it will not pay the firm to employ three units of factor, although the marginal product of the third unit is equal to the price of the factor, because the total product, which is equal to the sum of the three rectangles, is less than the total payment to the factor

which is three times the largest rectangle. It will pay to increase the use of the factor beyond three units so long as the marginal product of the quantity employed exceeds OC. This ceases to be true for OA units, and there the net product, i.e. the difference between the total product and the total payment

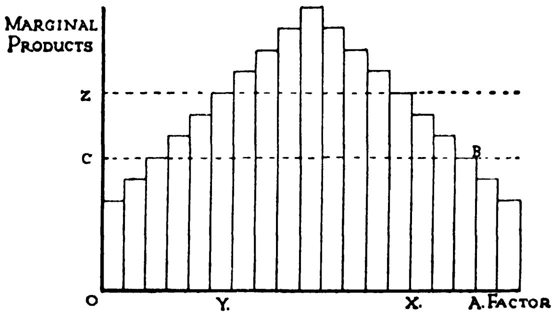


FIG. 19.

to the factor, is maximised. This point must necessarily be on the downward sloping portion of the diminishing marginal product curve. There will be some price OZ of the factor which is equal to the marginal product of OY units on the ascending side and of OX units on the descending side. This price OZ will be such that the total product of units O to Y falls short of $OZ \times OY$, the payment to them, by the same amount as the total product of the units Y to X exceeds the payment to them of $OZ\ (OX - OY)$. The employment of OX units would therefore result in a total product just sufficient to remunerate the variable factor employed. There would therefore be no advantage in employing OX units of factor or, what is the same thing, in paying a price OZ for the factor, but at any price below OZ it will pay to employ a quantity of factor greater than OX.

For some such quantity as OA, the marginal output AB multiplied by the number of units employed, or $OABC$ in the diagram, is less than the total of the products of all units O to A. $OABC$ represents the total payment to the factor in question, and the difference between it and the sum of the marginal products is the payment to the other factors of production.

The demand price, in terms of product, for a quantity of factor is therefore its marginal productivity; and the demand curve of an individual firm for a factor is part of the downward sloping portion of its marginal productivity curve for that factor. Since the firm is selling at a known market price, the demand schedule, in terms of money, is the physical schedule multiplied by the constant market price. Marginal productivity, which is a physical quantity, is not affected by changes in the price of the product, but the value of the marginal product, which is equal to marginal productivity multiplied by the market-price of the product, must, of course, vary with it.

A firm's demand curve for a factor of production consequently moves up and down with changes in the price of its product.

The aggregate demand curve for a factor of production is therefore obtained by adding the downward sloping portions of the curves of values of marginal products in all the uses to which a factor can be put.

Objections to Marginal Productivity Analysis

It might appear that, since the total product is represented as the total incremental products of units of one factor, the whole product should be paid to that one factor. It must be remembered, however, that any one of the variable factors—all of them if all factors are variable—can be treated in the same way, so that it is necessary to argue that each factor has a claim to the whole product. In fact, though it is convenient to talk of an added unit *producing* the increment in product, it is inaccurate and misleading to do so. The product is produced by the factors in combination, and the gross product of the whole is increased when the quantity of one factor is increased. The assumption that the quantity of all factors except one remain fixed is a device which becomes superfluous if mathematical methods are used.

An objection to marginal productivity analysis which is sometimes heard is that it depends on manipulation of arithmetical examples, but again these are a concession to the average reader of economics who in the past has usually been far from being a mathematician. Arithmetical examples prove

nothing, but they are useful for illustrating mathematical processes to those who cannot follow the mathematics.

A more serious misunderstanding, which has formed the basis of criticism of the concept by authoritative writers, is the belief that the addition of one unit of factor will not increase the product, but that the subtraction of one unit will throw units of other factors idle and decrease output by a greater amount. It is denied that the proportions in which factors are used can normally be varied. If a machine requires one man to work it, the withdrawal of a man means a machine is thrown idle also. First it is quite possible to take an entirely contrary view of the possibilities of variation of proportions in which factors can be used, but we will admit that there are cases where a rigidity of the one man one machine type exists, though we will not admit that it is a predominant feature of industrial organisation.

There is, however, a more conclusive reply. The marginal productivity of capital in a railway company is not to be discovered by imagining what the drop in receipts would be if a few miles of track were torn up or a bridge removed or a station pulled down, and the marginal productivity of capital in an engineering works is not the difference in output that would be made by adding a few lathes which would lie idle or a conveyor which would be superfluous. The marginal productivity of capital in an enterprise is the difference in the outputs of the existing concern and that of a somewhat similar concern in which equipment has been made just sufficiently cheaper to effect the saving of the dose of capital in question. The comparison is between two concerns with slightly different organisations of factors and not between the concern as it is and as it would be if deprived of a small but vital part of its organisation.

The difficulty can be traced to the problems of defining capital. If we talk of the marginal productivity of particular producers' goods, we inevitably involve ourselves in problems where labourers have no spades, power-stations are bereft of a dynamo and machines are idle because a motor has been taken away. If, on the other hand, we regard capital as a continuing fund of investment which happens to have been invested in a particular collection of producers' goods at the time in question, this difficulty does not arise.

A further objection to marginal productivity analysis was the alleged possibility that, if each factor was paid its marginal product per unit so that the total payment to it would be Marginal Product × Number of Units Employed, the sum of all such payments might be greater or less than the product itself. It is quite easy to show for a simplified case that the sum of the payments to factors at marginal productivity rates will exhaust the whole product. Let the supply of all factors to a line of production be variable, as indeed it must be in the long run, and assume that any factor can act as hirer of the others, so that the factors share out the product between them in some fashion or other. Then every *hired* factor must get its marginal product; for if it gets less, the demand for it will extend and its price will be bid up; if it gets more, the demand for it will contract. Again, every *unhired* factor must get its marginal product too; if it gets less, its owner will give up hiring other factors and organising them, and will hire his factor out to somebody else. If, on the other hand, it got more than its marginal product, owners of other factors would cease to let out their property and become hirers of factors. There will thus be a tendency to transfer factors from the hired to the unhired class and vice versa, until any excess of earnings over value of marginal product has been wiped out by competitive bidding.

Supply of Factor to Particular Occupations

We discovered the principle on which a given supply of factor is distributed when we saw that the price paid to a unit of factor in one occupation tends to be the same as that unit could earn in any other line of production.

Our study of the problem of marginal productivity has amplified that conclusion. An increase in demand for a commodity, causing its price to rise, increases the value of the marginal product of all variable factors employed in its production. A new equilibrium is set up, when factors have moved to equalise the values of marginal products in all its uses. The supply of a factor at a given price is thus determined by the demand for it in other lines of production, and the Total Demand Analysis which we worked out for the case of a stock of irreplaceable goods is applicable in this case.

The payments to identical units of factor will be the same in

all occupations only if the units of factor are free to move and their owners wish them to move. Acres of Wiltshire downland cannot be moved to the City of London to earn higher rents; it is indeed better to regard the two kinds of land as being different factors of production. In the case of labour, wage differences may exist between members of the same trade or profession in different parts of the same country, because people cannot afford the capital outlay involved in the change, or because custom and habit, or fear of being stranded in a community of strangers, keeps them from making a change. Movement across national frontiers is even more difficult, and in late years the difficulty has been generally increased by restriction on migration. Personal tastes for experiences of different types play a large part in directing choice of occupation. Some occupations give opportunities for foreign travel, others afford a good deal of leisure, others give security against the risk of unemployment, while some types of work are commonly considered to be more interesting or to carry more social prestige than others. Thus, in addition to some occupations being closed to many people, preferences for occupations vary.

From this latter cause we should expect that the tendency to equality of reward would not be towards equality of money payment, but towards equality of net advantages, including the non-monetary advantages of various occupations. We might expect that particularly interesting occupations which offer an unusual amount of leisure, a high degree of security of tenure or other particular advantages, would attract more people at a given wage level than would less attractive occupations. In such cases, the value of the marginal product of the people employed will tend to be less than that of similar people employed in other occupations which are less attractive. If employers who offer more attractive conditions can, in consequence, buy labour more cheaply, they will use it more extensively than those who, because of less favourable conditions, have to pay higher wages.

Although capital might be expected to be less discriminating in the matter of occupation, yet all the differences in remuneration cannot be traced to differences in the degrees of risk attaching to investments. At one time in England the social prestige and political advantage attached to landowning was

so great that land values were bid up to levels which did not permit a rate of return comparable to what could have been secured elsewhere with equal security. We shall, however, be more concerned with these causes of divergences from marginal product levels of remuneration, when we come to examine the conditions of supply for each factor separately.

Total Supply of a Factor

Writers of the Austrian school and American economists who have been greatly influenced by them have usually been content to assume the total supply of a factor as given, and to concentrate on the distribution of a fixed supply among rival uses for it. The English economists of the nineteenth century concentrated attention on the causes determining the total supply of a factor of production, and missed valuable clues which the other approach would have given them if they had followed it up simultaneously.

When the problem is approached by considering the division of a fixed supply of a factor among its different uses, it is difficult to miss the idea of the dependence of the reserve price on the demands for the factor in the alternative uses. It is not so obvious, unless one has that clue supplied, that the owner of a factor of production is also faced with the alternatives of using it or not using it. The supply of labour is dependent on two factors, the number of workers, i.e. the size of the able-bodied population, and the amount of work done per head. The first factor seems to follow no ascertainable rules, so that we must start with the amount of work done by a population of a given size. From the point of view of the individual worker, a further hour spent at work means an hour less of leisure, and he will try to strike a balance in just the same way as if he were exchanging one commodity for another. If free to choose the length of his working day, he will weigh the extra goods and services he can obtain with successive increments to his income against the corresponding cuts into leisure which increasing income entails. Equilibrium will be found when the individual is indifferent between the sacrifice of an extra hour of leisure and the earnings another hour of work will bring him. The marginal rate of substitution of leisure for income is equal to the wage rate.

Similarly for land, we can explain the total quantity of land to be let for rent in terms of the demand for income and the demand for land to use for the owners' pleasure. The fact that some landlords will let all their land makes no more difference to the method of analysis than would the discovery that Omar's vintner was a total abstainer. The one fact means that the demand for land is rather less than it might be (since certain landlords have no demand to use their own property) and in the other case the demand for wine is rather less than it would be if the vintner had a taste for his own wares. Alternatively, we can say that the market supply is greater than it otherwise would be. In the language of Chapter V, the price line representing the market level of rent is not tangential to any of these persons' land-income indifference curves.

Similarly for capital, we shall have equilibrium between the demand to obtain income by lending our money and the demand to have control of liquid resources.

Appendix

Real Costs

As we have seen, the utility school of economists explained demand for goods to consume in terms of a psychological concept, utility; a measurement of the satisfaction to be derived from consumption. They then required a quantity which should be the opposite of utility in order to explain equilibrium. This they found in the dissatisfaction incurred in the productive process and termed it disutility. Work, they said, when carried beyond a certain point becomes unpleasant; the longer the working-day, the more unpleasant is the last hour of labour. The individual will find equilibrium when the dissatisfaction arising from a further unit of work performed is only just compensated by the utility of the goods and services which the extra earnings represent. The marginal utility of income is then equal to the marginal disutility of work. This appears perfectly reasonable, until we realise that we are dealing with incommensurable quantities which can have no place in a scientific study. Having already rejected utility as the basis for the explanation of price, we cannot retain disutility.

A similar explanation of the conditions of supply of capital discussed an equilibrium between the satisfactions arising from

the provision of income in the future and the dissatisfactions arising from the abstinence required to save the capital which should provide the future income. The payments for labour and capital were thus to be explained as the rewards necessary to induce people to incur these real costs of production. In the case of land, however, there was no such item of real cost corresponding to the unpleasantness of labour or abstinence. If the labourer or the saver of capital were not adequately remunerated, he would refuse to incur the dissatisfaction of providing his factor of production, but land would be available in any case. Hence it was argued that the rent of land does not enter into the cost of production.

The argument can be refuted in detail or on general grounds. It is hardly possible, seriously, to contemplate a Rothschild or a Rockefeller being rewarded by the receipt of interest for having exercised the heroic virtue of thrift in abstaining from consuming his fortune in riotous living. It is a serious defect in the disutility hypothesis that it fails to give an explanation of the rent of land. The method of displaced alternatives enables us to handle all factors of production and, indeed, all goods and services in a uniform fashion. It would therefore be preferred to the disutility hypothesis even if the latter were intrinsically equally plausible, which we have seen is not the case.

Writers who followed the Austrian method spoke of real cost of production as displaced alternatives. Thus the real cost of using a ton of coal to drive a railway locomotive would be the steel it might have produced, or the amount of sea transport it might have given if burned in a ship's stokehole, or the heat it might have given in the domestic grate. The real cost of using a man's labour in digging a railway cutting would be the amount of wheat he might have produced if put to cultivate land instead. Because the real cost of using a factor was looked upon as being equal to any one of the alternative products forgone when the decision was made to allot it to one particular purpose, the terms Displacement Cost and Opportunity Cost have been used to describe the concept. The displaced alternatives include the uses the owner of the factor can make of his own property, so that when we are thinking of the supply of some kind of labour as a whole, we need only

consider the two possibilities of using time for earning money income and using it for leisure. Similarly with capital, it is a question of lending versus hoarding.

The Austrian definition of real cost has in turn been attacked as unnecessarily circular, and in more recent work real cost of production is defined in concrete terms as the physical quantities of factors required to make a unit of commodity, and the definition goes on to state that these have a price because they have alternative uses.

Further Reading

Black, J. D.: *Production Economics*, Part 3.
Boulding, K. E.: *Economic Analysis*, Chaps. 9, 10 and 11.
Clark, J. B.: *Distribution of Wealth*, Chaps. 11–24.
Knight, F. H.: *Risk, Uncertainty and Profit*, Chaps. 4 and 5.

Chapter IX

OUTPUT IN A COMPETITIVE MARKET

The construction of a supply curve for a commodity in current production, which we undertook in Chapter VII, was based on two assumptions about the market. First, we assumed that our producer could take market price as given, and that no manipulations of the size of his output would affect the price in the market. Secondly, we assumed that our producer was able to buy factors of production on the same terms as any other producer catering for the same market. Both these assumptions require amplification, and the first step is to decide exactly what we mean by a commodity.

What is a Commodity?

We have been using the word commodity to indicate a class of objects so closely identical that a consumer would be quite indifferent which particular units he secured. If we take wheat as a commodity, one bushel of wheat should be a perfect substitute for any bushel of wheat, one bale of cotton for any other bale. The slightest acquaintance with commodity markets will show this to be untrue, for wheat and cotton prove to be the general names given to classes of product which, although they belong to the same botanical species, have quite different industrial uses. The soft English wheat is not, in the bakers' opinion, a perfect substitute for imported wheats of a higher gluten content. Still, there are recognised grades both of wheat and cotton, which are so narrowly defined that purchasers are willing to buy without first seeing what they are buying. But although a buyer will be indifferent between any two bales of fully middling American cotton in Liverpool, he will certainly not be indifferent between them if one is in Liverpool and the other in New Orleans. They are not then

perfect substitutes one for the other, and he will not, in general, be willing to pay the same price for them.

To turn to finished goods purchased by the general public, it is possible to sell a simple chemical substance under a trade name, to advertise it until consumers believe it possesses unique properties, and to charge a price many times that at which the substance can be bought under its true name. Two pots of jam boiled together may, by acquiring different labels, be subject to entirely different conditions of demand and be sold at different prices. They are no longer perfect substitutes for one another. Not only is one brand of cigarette not a perfect substitute for another brand made from the same type of tobacco and sold at the same price, but a packet bought at one shop may not be a perfect substitute for a packet bought at another in the same street. Convenience of access, admiration for the girl behind the counter, friendship for the proprietor and a score of other motives attach certain purchasers to certain shops. Preferences based on small or imaginary differences between products lead consumers to purchase one brand exclusively.

The definition of a commodity which emerges from this discussion is that it consists of a group of products which are sufficiently close substitutes to make it convenient to class them together. The differences in preferences which prevent their being perfect substitutes may be due to technical differences, differences in location, or no ponderable differences may exist, but so long as consumers do not rate them equally on their preference scales, they will not be perfect substitutes.

For Pure Competition a Commodity must be Highly Standardised

We have built up our supply curve on the assumption that the producer can sell as many units as he chooses at the market price, and this condition can only be true if his product is a perfect substitute for the products of all other producers. The fact that he can sell any output he chooses at market price means that the demand curve which faces an individual producer is a horizontal straight line. This, in turn, means that elasticity of demand is infinite; for on a horizontal demand curve, the proportionate change in price, which corresponds to

a finite proportionate change in quantity, is zero, and the value of a fraction with a denominator of zero is infinity.

We saw that the degree of elasticity of demand for a commodity depends on the ease with which it could be substituted for other commodities. Consequently, there can be but one price in the market. It will not pay any one producer to undersell his fellows, because he can sell the whole of his output in any case. If he tries to charge more than his fellows, his sales will drop to zero.

The Number of Competitors must be Very Large

It is improbable that perfect substitution would continue to exist for very long if the number of producers were limited, but it is certain that the condition of a single price need not exist. If the output of one producer represents such a large proportion of market supply that its subtraction will affect price, then that producer can manipulate the market price by adjusting his supply. If sales are made by a process of competitive bidding by buyers, one very large seller will be able to raise the market price by decreasing his supply. He will have to be a very large part of the market if he is to reap any benefit from the manœuvre, otherwise he will merely be enabling his rivals to sell larger quantities at the old price.

If, however, it is the custom for a seller to set a price which the buyers can either take or leave, it may be possible for one seller to ask a price slightly higher than his rivals, and if his share of the market is too large for his rivals to replace him by expanding their own outputs, a certain number of buyers will have to pay that price or go without.

A state of *pure competition* is said to exist when two conditions are fulfilled:

(1) The product of one producer is a perfect substitute for the product of any other producer.

(2) The number of producers is too large for any one of them to influence market price by withholding his output.

Perfect Competition and Factor Supply

Certain writers have made a distinction between a state of pure competition and one of perfect competition in which, in addition to the conditions laid down above, there is the further

proviso that all rival producers shall be able to buy factors of production on identical terms. An alternative statement of this third condition is that the factors used by all are identical, and the number of producers so large that the demand for a factor from any one of them is too small a part of the total demand for the factor to influence its market price.

The Output of an Individual Firm in a Competitive Market

We must now examine more closely the conditions which determine the contribution by an individual producer to market supply. We shall retain the condition that each firm has a fixed quantity of at least one factor of production, so that limits are set by the Law of Diminishing Returns to the range of outputs it can produce. We shall no longer confine ourselves to the problem of the determination of the output a particular firm will produce at a given market price, but will also examine the question of the price level which will enable a firm to continue to exist in the long run, and thus we shall be able to say something about the conditions affecting the number of firms in the market. First, however, we must return to our examination of the marginal cost scale, and now we must allow for the existence of fixed cost items in addition to variable cost items.

Average Total Costs and Marginal Costs

In Table 4, on p. 86, we showed the relation between average variable cost and marginal cost. We will now recalculate these costs on the assumption that in addition to variable cost there is a fixed cost item of £17·6.

Marginal cost has remained unchanged, as we should expect, since it is dependent on variable cost. The marginal cost of, say, the seventh unit is still 2·4, but it is now the difference between 46·5 and 44·1 instead of between 28·9 and 26·5. The scale of average total cost lies above the scale of average variable cost, but the difference becomes progressively smaller as the fixed cost is divided between an increasing number of units. The minimum value of average cost now occurs for a larger number of units than before, i.e. for **12** and **13** units instead of **11** and **12**, as the increasing cost of variable factors is temporarily offset by the decreasing cost per unit of the fixed cost item. The

larger the proportion of fixed cost in relation to variable cost, the greater will be this shift in the point of minimum average cost. Again we see that the average and marginal cost scales cut at approximately the point where average cost is a minimum.

TABLE 5

(1) *Units of Output*	(2) *Total Cost* £	(3) *Marginal Cost* £	(4) *Average Cost* (2) ÷ (1)
0	17·6	—	—
1	25·1	7·5	25·1
2	30·6	5·5	15·3
3	35·1	4·5	11·7
4	38·5	3·4	9·6
5	41·4	2·9	8·3
6	44·1	2·7	7·4
7	46·5	2·4	6·6
8	48·8	2·3	6·1
9	50·4	1·6	5·6
10	52·0	1·6	5·2
11	54·3	2·3	4·9
12	57·6	3·3	4·8
13	62·4	4·8	4·8
14	68·6	6·2	4·9
15	77·6	9·0	5·2

In numerical examples it is inevitable that there shall be two equal values of average cost at the minimum which, strictly, lies between them. Only if infinitesimal steps are taken is there a single-point minimum, and then the numbers become so complicated that the objective of numerical examples—simplicity of exposition—is lost. To determine precisely the relation between average and marginal costs, we shall have to have recourse to graphical methods.

Marginal Cost Redefined

Before we proceed with the determination of the output of a firm by graphical methods, we must return to the definition of marginal cost, which for the purposes of numerical examples we defined as the extra cost of adding an extra unit of output. Marginal cost being the increase in cost for an increase of one unit in output is thus the rate at which cost is increasing per unit of output, but it is quite apparent from the Total Cost Curve in Fig. 20 that the rate of increase in cost at any given

point on the curve depends on the size of the unit of output. At some points on the curve, the rate of increase is smaller for a larger unit than for a smaller one, i.e. where the curve is convex upward; where it is concave upward the reverse is the case. We can get over this difficulty if we imagine the increase in output to be of infinitesimal size. Then, if we have an output of x units already being produced at a total cost of y and a very small increase in output Δx increases cost by Δy, marginal cost is equal to $\frac{\Delta y}{\Delta x}$. Our definition of marginal cost is then:

The marginal cost of x units of product is the rate at which total cost increases for a very small increase in output when x units are already being produced.

To represent this graphically, let the curve CC of Fig. 20 represent the relation between total cost and output. The curve starts at a point C on the cost axis, whose height OC represents the fixed costs which must be incurred if the firm is to remain in existence even if no output is produced. The shape of the curve depends on variable costs. An output x can be produced for a cost y, and the addition of a small quantity of output ab increases costs by bc. Marginal cost of x units is equal to bc/ab or the tangent of the angle bac or the slope of the chord to the curve between the points a and c.

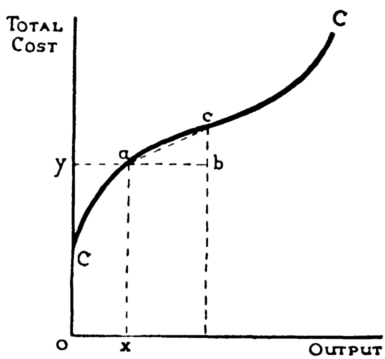


FIG. 20.

Now, if the increase in output ab is very small indeed, so that the points a and c tend to coincide, the curve ac and the chord ac tend also to coincide, and the slope of the chord becomes the slope of the curve at the point a. Therefore we can represent the marginal cost of a given output by the slope of the total cost curve at the point on it corresponding to the given output.

If we travel along the cost curve from left to right—that is, in the direction of increasing output—we see that the slope of the curve decreases at first, reaches a minimum, and then increases. Since the slope of the total cost curve represents marginal cost, we have again arrived at the U shape of the marginal cost curve which we observed in our arithmetical examples. This U shape is a logical consequence of the Law of Diminishing Returns. We saw that this law gave us a total product curve of a particular shape, which in turn, when replotted, gave us a total cost curve of the shape we have employed. We have now shown that such a total cost curve must have associated with it a U shaped marginal cost curve. Let us now see how we can represent average cost.

Graphical Representation of Average Cost

In Fig. 21, C_1C_2 represents the relation between total cost and output.

The total cost of producing output OX is PX, and the average cost per unit is $\frac{PX}{OX}$, which is the tangent of the angle POX or the slope of the line OP.

Now let us imagine point P as being attached by a piece of elastic thread to O, and moving along the curve from C_1 to C_2. As P moves upward and to the right, the line OP rotates in a clockwise direction, and its slope decreases until it comes into the position OQ corresponding to output OM. After passing Q the line rotates in an anti-clockwise direction and its slope increases. Thus, average cost, like marginal cost, decreases at first, reaches a minimum, and then increases and its curve too is U shaped.

Further information can be extracted from this diagram.

For all values of X from zero to OM the slope of the line OP is greater than that of the curve, so average cost is greater than marginal cost.

At Q corresponding to output OM, the line is a tangent to the curve, and the slopes of the line and of the curve are the same. For this value of output, therefore, average and marginal costs are equal, and, as we have already seen, average cost is a minimum at this point.

For outputs greater than *OM*, the slope of the curve is always greater than the slope of the line *OP* and marginal cost is greater than average cost.

We have thus confirmed the picture of the relationship of the marginal and average cost scales which we obtained from the

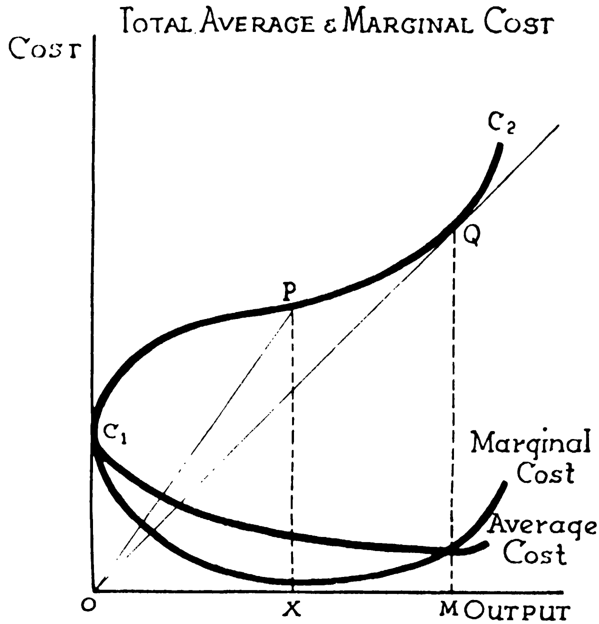


FIG. 21.

arithmetical examples, and we have proved our earlier conclusion that the two scales cross at the point where average cost is a minimum. For this size of output average cost and marginal cost are equal; for smaller outputs marginal cost is less than average cost, and for larger outputs marginal cost is greater than average cost.

Once more let us remind ourselves that these results are deductions from the Law of Diminishing Returns, on the truth of which our explanation of supply prices in terms of cost entirely depends.

Before we can use this information to discuss the output of a firm, we must examine the nature of the cost items we have encountered so far a little more closely.

Market Prices and Contingent Rewards of Factors

When we are costing the consumption of each factor per unit of output, we shall usually have no difficulty in finding a price at which to value it. We shall be able to find market prices for many of them, to obtain price quotations for material and stores, to ascertain trade-union wage rates for labour and so forth.

There will, nevertheless, be cases where we cannot discover a market price. Some factors will be working on terms which do not include a fixed rate of remuneration. They will work for what they can get, and what they do in fact get will depend on the success of the enterprise in which they are engaged.

Such factors are usually relatively immobile once they have become attached to one production organisation. The capital subscribed by ordinary shareholders may have been spent on objects which have little or no use for any other purpose. An extreme case is that of the investor in coal-mining whose capital has been spent on sinking a shaft and preparing a seam of coal for extraction. If no coal proves to be there or if it is of quality too poor to be worth extracting, the investment is entirely lost. The shaft is worth nothing for any other purpose. At the other extreme, machine tools which were purchased to make motor-car-engine components may be used to make vacuum-cleaners or washing-machines.

Between these extremes lie many cases where adaptation can take place with a greater or less degree of difficulty.

There is, however, one means by which the capital invested in fixed assets can be moved out of occupations which can no longer pay normal rates of return into occupations able to pay higher rates. Fixed assets wear out, and during their lifetime they should produce a depreciation fund adequate to replace themselves when they have to be scrapped, and over and above that an annual payment for their hire. If such an asset cannot earn an annual sum sufficient to cover both these items, depreciation will not be replaced. The asset will be allowed to wear out unreplaced, and the annual depreciation allowances will be invested in assets of another type.

Some forms of labour service will also be in receipt of contingent rewards. Payment for certain types of work is on a

commission basis, being a percentage on the volume of sales or on the amount of profits. The owner-manager of a one-man business or the members of a partnership may be in receipt of incomes which are dependent on the success of their activities, and which may be regarded as the reward of their personal enterprise rather than as payment for the employment of capital. To carry on the business of a broker, whether in the bill or insurance markets, the stock exchange or commodity markets, involves the employment of relatively little capital but a great deal of personal activity. A broker will have to pay out sums regularly for clerical labour, office rent and expenses, but the amount of his own remuneration depends on the difference between receipts and expenses, and will fluctuate with the volume of business. It is a contingent reward.

Money Expenses and Total Costs

We can divide the payments, which are made to factors of production, into two groups: the payments which must be made to all factors that are hired for an agreed money payment, and the contingent payments to those factors which claim any residue of net receipts and which are usually the hiring factors.

The money expenses include both payments to the variable factors which can be hired and fired, bought and sold at short notice, and also payments to the more immobile factors, which payments must be made if the firm is to continue in existence, even if it produces nothing.

The first of these will include wages of operative labour and payments for materials, stores, fuel and power; while the second will include wages of supervisory and managerial staff, rents of land and buildings, costs of repairs and maintenance, and interest on money borrowed at a predetermined rate. Over fairly short periods of time, i.e. periods not exceeding a few months, this second group can be neither increased nor decreased, so that over short periods of time they and the factors receiving contingent rewards will form the group of fixed factors which we require if we are to apply the Law of Diminishing Returns. To this fixed organisation we consider varying quantities of the variable factors to be applied.

Output of a Firm in the Long Run

We can now use what we have learned of the nature of costs to re-examine the problem of the size of the output of a firm and to throw further light on the total supply of a commodity. We have discovered two conditions determining the output of a commodity from a single firm, operating in a competitive market, where it can sell whatever quantity it chooses at the market price.

First, the output of a firm is that quantity whose marginal cost is equal to market price.

Secondly, the price at which an output of a given size is sold must be greater than the average variable cost of producing that output.

From this we conclude that the supply curve in a competitive market is that portion of a marginal cost curve which lies above the average variable cost curve.

Thus we can explain the total output from the firms in existence at a moment of time, but this analysis as yet tells us nothing about the number of firms, whether it is expanding, contracting or stationary. To answer this question, we must investigate the relationship between price and the output of a firm over a period of time long enough to allow the quantities of all factors of production employed to be varied. In Fig. 22 that portion of the Marginal Cost curve, *MC*, above the point where it crosses the Average Variable Cost curve, *AVC*, will be the short-run supply curve of the firm in question. We have now to decide at what point on this curve prices must settle if the firm is to be able to pay all the factors of production it employs as much as they could expect to earn elsewhere, and so to persist in the long run. The curve of Average Total Cost, *ATC*, includes not only the variable costs represented by *AVC*, but all payments to factors which must be made if the firm is to continue to exist.

Suppose that a firm finds itself selling an output Oq_1 at price Op_1. Total receipts are measured by the rectangle Oq_1ap_1, but total costs are equal to Oq_1bc. Receipts thus fall short of the amount necessary to pay all the factors their market prices and thus retain their services. Sooner or later the concern will shut down, if this state of affairs continues.

If this condition is common to a number of firms in the industry, the total supply of the commodity will diminish as firms close down, and there will be a tendency for price to rise.

If, on the other hand, a firm is producing output Oq_2 at price Op_2, total receipts equal Oq_2dp_2 and total costs Oq_2ef; there is

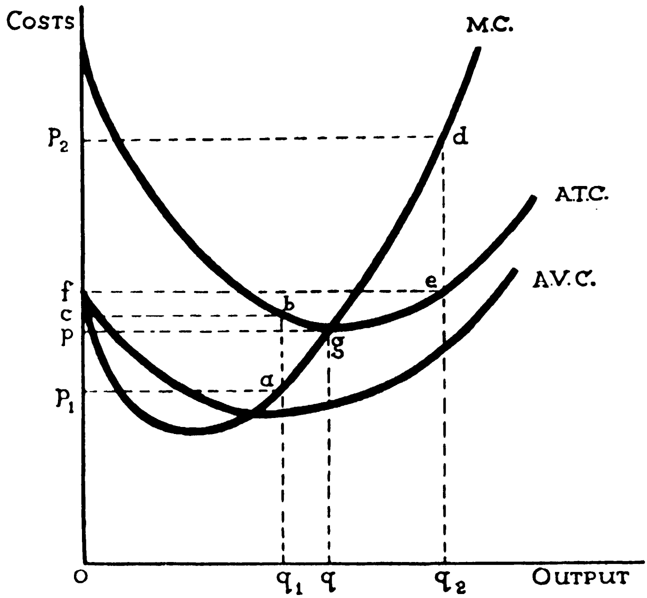


FIG. 22.

a surplus of receipts over and above the sum necessary to pay all factors their market prices. This surplus will be appropriated by the owners of the most immobile factors, for it will not be possible at short notice to detach supplies of these factors from other lines of production, where they are occupied, and to bring them in to compete with identical units, which are receiving a higher reward. In the longer run this movement will take place, if such excess payments exist on any important scale, and then, as new firms are set up, the supply of the commodity will be increased, price will fall and the surplus receipts will disappear. Price can therefore neither permanently fall short of average total cost, or permanently exceed this level, without setting up forces which tend to restore

the equality. At the same time, the original condition, that price is equal to marginal cost, must also hold, and there is only one point at which both these conditions can be fulfilled, that which makes average total cost a minimum. Long-run equilibrium is to be found therefore at price *Op*, which calls forth a supply whose marginal cost is equal to average cost.

It is clear, from what has gone before, that the marginal cost of each firm's contribution to an output will tend to be the same and equal to market price. This result comes about, not because businessmen use marginal cost accounting, for few of them have ever heard of it, but because, by trial and error, they are attempting to maximise the return to the fixed factors of production, for these are in general the ones they themselves own.

Under the conditions we have stated, minimum average cost will be the same for each of the firms in an industry, because we have supposed that they are purchasing some of their factors in the same market, and there is the tacit assumption that the fixed factors owned by the firms are of identical quality, though each firm does not possess the same quantity. In these circumstances minimum average cost will be the same for everybody, though outputs will not be of the same size.

We can get nearer to reality, without seriously impairing the conditions of perfect competition, by removing the assumption that factors of production are of homogeneous quality and, still in the long run, minimum average cost will be the same for all firms, but the full explanation of this is better reserved for a later occasion.

Joint Supply and Joint Demand

So far we have regarded a firm as producing a single product, but this is in fact a circumstance not commonly encountered. It is more usual to find the same factors of production being used to make more than one product, and it is usual to describe such products as being jointly supplied. Sometimes joint supply is undertaken because for some reason or other it is more convenient. Thus, in this country farming is of the type which is usually termed "mixed." The reason for this may be a desire to avoid the risks of total crop failure if only one crop is grown, by growing crops requiring different weather conditions,

so that if weather proves unsuitable for one, it may nevertheless suit another. A more important reason probably lies in the variety of types of land often found on a single farm in this country. It is more profitable to select crops to suit particular fields than to treat the whole in the same way. Similarly, there are cases of firms who are catering for a seasonal demand who find other products which they can make during the rest of the year and so prevent equipment and labour being idle. Then there are many cases where a process provides a product and one or more by-products. A gas-works produces coke and a number of chemical substances; sheep-farming provides mutton and wool, and so forth. In the majority of cases it is possible to vary the proportions in which the joint products are supplied, but in a few cases this is not so. The wheat-grower can vary the proportions of grain and straw by selecting the strain of wheat, but as yet it has not proved possible to vary the proportions of cotton-seed and lint. The proportion is not fixed, but it varies with weather conditions and the activities of the boll weevil, and not with the intentions of the planter. Where the proportions are variable, it is possible to calculate the marginal cost of each of the joint products by considering the quantity of one of them to increase, while that of the other remains fixed. The output of either can then be regulated to equalise price and marginal cost.

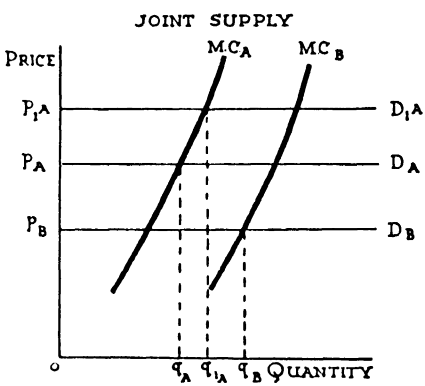


FIG. 23.

In Fig. 23, let MC_A be the marginal cost of increasing the output of A, leaving the output of B the same, and similarly let MC_B be the curve of marginal cost of B when the output of A is constant. Then, if the market price of A is p_A and that of B is p_B, the demand curve for each product of the firm will be a horizontal straight line at the appropriate height. The points of intersection of marginal

cost and demand curves will give the contributions the firm will make to market supply. Then let the price of *A* be raised to p_{1A}; output of *A* will be increased until the marginal cost curve cuts the raised demand curve, while the supply of *B* will not vary. In a similar way we can deal with a decrease in the demand for a product.

The case where the proportions of the products are fixed cannot be treated in this way. Since it is impossible to increase the output of either separately, we cannot determine the marginal cost of either. We do know the marginal cost of increasing the output of both together, and we can say that output of the two together will be increased, so long as the increment in cost incurred in producing a further "unit" of the combined products is less than the increment in receipts from selling it when produced. This increment in receipts needs some explanation. If the price of cotton, say, rises and the output of cotton is increased, the output of cotton-seed is increased. The demand for cotton-seed has, however, remained unchanged, so that its price falls. The increment in receipts from the whole price movement is therefore composed of two items: (*a*) the change in receipts due to selling a slightly larger quantity of cotton at a higher price, and (*b*) the change arising from selling a slightly larger quantity of cotton-seed at a lower price. The latter item may be positive, negative or zero, according to the elasticity of demand for cotton-seed.

Rival and Complementary Supplies

These two cases—the first where the proportions in which the joint products are produced are completely variable, and the other where these are rigidly fixed—by no means exhaust all the possibilities. They are indeed particular cases of two possible types of relationship. Just as we saw that some goods were rivals in consumption so that an increase in the consumption of one led to a decrease in consumption of others, so we have goods which are rival in supply, in that an increase in the supply of one decreases the supply of others. It does this because the supply of each of a group of rivals is dependent on the same factors of production, and if one secures a greater supply of factors, the others have to go short. The tendency to rivalry in supply will not be so general as the tendency to rivalry in

demand, because there is to some extent specialisation of factors of production, and it is where production is dependent on a specialised factor that rivalry will be most intense. Thus, all the commodities produced in a given community are in some degree dependent on the same pool of unskilled labour, and are therefore to some extent rivals, but we should not expect the degree of rivalry to be very intense. Wheat and sugar-beet, on the other hand, are rivals for the same type of land, and the rivalry between them will be more intense.

Again, just as goods are complementary in demand and an increase in demand for one increases the demand for the complements, so a complementarity relationship can exist in production. The cotton-seed/cotton-lint example is a case of complementarity of a type that is no doubt extremely common, but it is quite possible for production of the complement to increase less than proportionately or more than proportionately to the change in output of the commodity in which we are interested.

Effects of Rivalry and Complementarity on Prices

If the production of B is rival to that of A, then an increase in the demand for A, causing all producers to move up their marginal cost curves, will attract away factors which are used to make B, and the supply of it *at all prices* will be diminished. So in Fig. 24A, S_A represents the market supply curve for A, and quantity q_1 is being produced at price p_{A1}. Demand for A then increases, so that price rises to p_{A2} and supply increases to q_2. The factors to produce this extra quantity are wholly or partially withdrawn from the production of B, so less of B will now be produced at all prices and the market supply curve for B is moved to the left from S_{B1} to S_{B2}. The market demand for B, as represented by D_B, has remained the same, and as the supply curve for B has moved to the left, S_{B2} will cut the demand curve at a higher point and the price of B will be p_{B2} instead of p_{B1}. When two commodities are rival in supply, therefore, an increase in demand for one of them will lead to a rise in price of the other, and, vice versa, a fall in the demand for one will lead to a fall in the price of the other.

Similarly, if B is a complement to A in production, and there is an increase in demand for A so that its price rises from p_{A1} to p_{A2} and production is increased according to the

supply curve, the supply of B will be automatically increased. More of B will be supplied at all prices, so the supply curve

RIVALS IN PRODUCTION

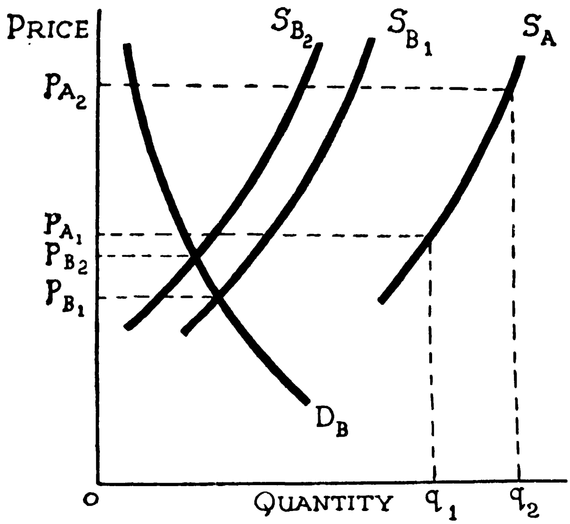


FIG. 24A.

for B will move to the right from S_{B1} to S_{B2} (Fig. 24B). Demand, however, will continue to be represented by D_B, so that as the supply curve has moved to the right, it will now cut the demand curve at a lower price than formerly, and the price of B will fall from p_{B1} to p_{B2}.

When the relationship of two or more goods is one of complementarity in production, an increase in the demand for one will lead to an increased supply and a lower price for its complements, and a decrease in demand will lead to a decreased supply and a higher price for its complements.

Joint Demand

When, in our study of the Theory of Choice, we observed the possibility of a relation between the demands for two or more commodities, by which an increase in demand for one led to an increase in demand for the others, we were concerned with cases of related tastes; but this does not exhaust the study of

complementarity in demand. There is a joint demand for factors of production to make a product. We have seen from our discussion of the Law of Diminishing Returns that it is possible technically to use a wide variety of combinations of proportions and the choice of one particular combination depends

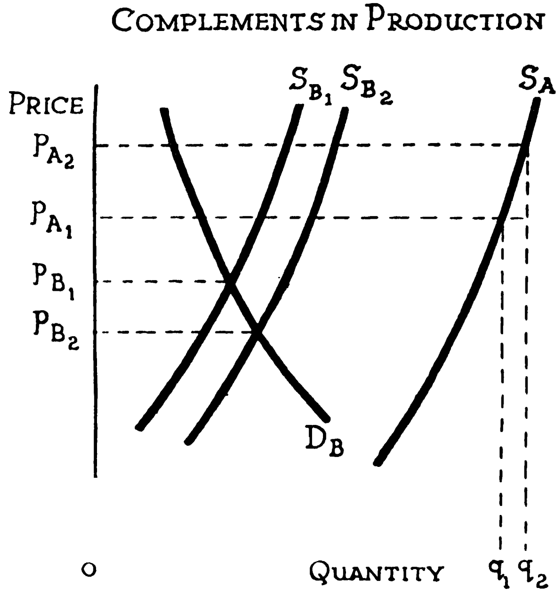


FIG. 24B.

on the relative prices of the factors. In the general case, one factor can be substituted for another if price changes make it profitable to do so. In most manufacturing processes it is possible to use more machinery and less labour, or vice versa, as the price of each decides. In farming, more or less intensive methods can be used—that is, the quantity of labour to an acre of land can vary within wide limits. If one factor of production is relatively dear, the effort will be made to use as little of it as possible to get the maximum product out of every unit. This will be done by using greater quantities of the cheaper factors. In Canada, where land is cheap relative to labour, land is cultivated "extensively," and labour-saving methods, which result in low yields per acre but high yields per man, are employed. In England, where land is relatively dear,

the effort is made to get higher yields per acre. The effect of a rise in the price of one factor due, say, to a demand for it elsewhere, is to increase the demand for other factors, which will now be used in greater quantity and whose price will, in consequence, tend to rise.

It may, however, be the case that substitution is impossible: that factors must be used in fixed quantities. Something approaching this is to be found in the cutlery trade, where machines which can take account of slight differences in the quality of steel have not been invented, and the forging and grinding of better-class cutting tools has to be done by hand. In the short run we may find it impossible to interchange labour and machinery in almost any industry. We also get cases of products made up from components like knife-blades and handles, golf-clubs and shafts. If the price of one factor or component rises, another factor or component cannot be substituted for it, and the price of the finished article must rise. This will cause contraction in the demand for it, and in the demand for the other factors or components; hence their prices will fall.

An exception to this will occur when the commodity whose price rises represents a very small part of the total cost of the final product. If this is so, the increment to the cost of the final product due to the price change may be less than the minimum sensible. Thus the cost of sewing cotton in the garments for which it is used, and of manganese or certain rare earths in the cost of special steels, is such a small item that substantial percentage increases have an effect on the total cost which is hardly appreciable. The proportion of labour cost in total cost of the product is highly important from the point of view of trade-union policy.

FURTHER READING

BOULDING, K. E.: *Economic Analysis*, Chaps. 5–8.
KNIGHT, F. H.: *Risk, Uncertainty and Profit*, Chap. 6.
ROBINSON, E. A. G.: *Structure of Competitive Industry.*

MORE ADVANCED READING

CHAMBERLIN, E.: *Theory of Monopolistic Competition*, Chap. 2.
RATZLAFF, C. J.: *Theory of Free Competition.*
ROBINSON, J.: *Imperfect Competition*, Book 3.
STIGLER, G. J.: *Theory of Price*, Chaps. 9 and 10.

Chapter X

OUTPUT IN A MONOPOLISTIC MARKET

We have discussed the problems of determining outputs and prices when the number of producers in the market is very large. This is the problem of pure competition. We must now examine the question of the output of a firm when the number of firms competing is small enough for the effects of changes in the output of one to affect the prices at which the others can sell their outputs.

We shall find that the same principle which enabled us to determine output in a competitive market—that each producer will fix his output at the level most profitable to himself—will prove just as useful here. Just as before, each producer will have a certain total of expenses which he can avoid by refraining from producing an output of a given size, the amount of the expenses being related to the size of the output. This total we have called the variable cost of the output in question. It will be the purpose of the producer to make the difference between total receipts and total variable cost a maximum, and it is with this end in view that he will decide on the size of his output. The analysis will be a little more complicated than it was before, for in a competitive market we can assume that a producer can always sell one more unit at the market price, so that receipts per unit remain constant while costs per unit change.

When we introduce the idea of monopoly, we can no longer make this assumption about price, so that both receipts per unit and cost per unit will be variable.

What is a Monopolistic Market?

As we have already seen, if we remove the assumption that the number of firms is too large for any one of them to have the power to influence market price, we can no longer rely on

the product of any one firm being a perfect substitute for the product of any other. It was the existence of perfect substitutes for the product of an individual firm which made elasticity of demand for it infinite and the demand curve a horizontal, straight line. Remove this condition of perfect substitutes for a firm's product and the demand curve for that product will slope downward to the right.

We can picture this occurring because each firm has attached to it a number of customers who will buy from it rather than from any other seller asking the same price. Thus, no firm will be able to increase its sales by merely offering more for sale at the same price in the way a firm in a competitive market can do. Once a firm in a monopolistic market has satisfied the demand of the customers who normally buy from it, the only method to increase sales is to attract away some of the customers attached to its rivals. Ignoring for the present the effects of advertising, the only way a firm can do this is to cut price.

Consequently, an increase in sales, whether to its own clientele or to people who are buying from rivals, can thus be secured only by lowering price, and the demand curve relating to the product of a single firm slopes downward from left to right. We can, if we like, think of each firm as working in a sub-market of its own, whereas in competitive conditions there is one unified market for everybody.

The popular view that monopoly power lies in control over supply is not a truly accurate account of the matter, for every producer has control over his own supply. Monopoly power lies essentially on the side of demand. Some element of it exists whenever the product of one producer is not a perfect substitute for the products of other producers.

A monopolistic market can therefore be defined as one in which there is no perfect substitute for the product of an individual seller, so that there is a separate demand curve for the product of each seller in the market.

It is clear from this definition that we shall find varying degrees of monopoly power, its effectiveness in any one case depending on the remoteness of substitutes. If there are many very close substitutes for the product of a firm, the conditions of demand for it will differ but little from those of pure compe-

tition, but if the product is not readily substituted by or for others, a considerable fall in price may lead to little substitution for other products and a rise in price may have little effect in causing other products to be substituted for it. It would appear, therefore, that there is a close connection between elasticity of demand for a product and the power of the producer of that product to pursue a monopolistic policy.

Elasticity of Demand and Monopoly Revenue

In order to examine the basis of monopoly policy, let us look once again at the simple case of a fish market with a catch to be disposed of and no demand on the part of the sellers for their own product. The catch is one of 15,000 lb. of fish, and under competitive conditions the market would be cleared at a price of 4*d*. per pound. Next we suppose that a ring is formed to dispose of the catch on more advantageous terms. Since the catch is in existence, the policy of the ring will be to obtain the greatest total return, which is the same thing as the greatest total outlay on the part of the consumers. This will be achieved by fixing the reserve price at 5*d*., at which figure the market will be willing to take 14,000 lb. for a total outlay of 70,000*d*. The ring could have obtained the same result by throwing 1,000 lb. back into the sea, reducing supply to 14,000 lb., and leaving the market to fix the price at 5*d*. In either case the 1,000 lb. will be thrown away. A monopoly can always operate by fixing its reserve price and leaving buyers to decide what is sold, or by fixing supply and leaving buyers to fix the price. The choice of method will depend on which is most convenient in the circumstances of the market, but the result will be the same in either case.

TABLE 6

Price	*Buyers' Demand*	*Consumers' Outlay*
	lb.	*d.*
8*d*.	7,000	56,000
7*d*.	9,000	63,000
6*d*.	11,000	66,000
5*d*.	14,000	70,000
4*d*.	15,000	60,000
3*d*.	18,000	54,000

An examination of the demand schedule will show that, at prices below *5d.*, consumers' outlay declines for a fall in price, and that, for prices above *5d.*, outlay increases for a fall in price. *Demand is therefore inelastic below 5d. and elastic for prices above 5d.* A policy of restriction of supply, or of price raising, increases receipts only so long as demand does not become elastic. This agrees with what we know about the relation between elasticity and substitution, for the higher the price of a commodity goes the more commodities are likely to become close substitutes for it.

Monopoly Policy with Costs of Production

The policy of attempting to make total receipts a maximum is applicable only in the case of a stock of commodity which is already in existence. It is not appropriate when a decision as to how much to produce has to be made, for the monopolist will not be interested in maximising the return to all the factors of production employed, but only in securing the greatest possible return to those which he himself owns. Factors he hires or purchases will be paid their market prices and no more. The sum of payments made to factors of production we will call the monopolist's "money expenses of production." This will not correspond to "prime costs," because it may include such items as rent, which the monopolist has contracted to pay over periods of years and which is an item of his fixed cost. When he has paid these money expenses, he appropriates the balance of net receipts as the payment for the factors he owns, and it is this amount which he will try to maximise.

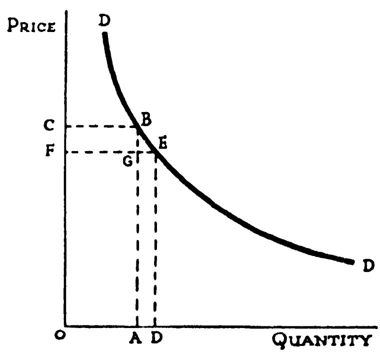


FIG. 25.

Marginal Revenue Defined

A monopolist who is varying his output in an attempt to arrive by trial and error at the greatest possible net receipts will be watching changes in two quantities. He will watch how total cost changes with output,

that is, he watches marginal cost; and he will also watch changes in total receipts, but as yet we have given no name to the latter. In Fig. 25 the rectangle *OABC* represents total receipts from selling a quantity *OA* at price *OC*. Rectangle *ODEF* represents the total receipts obtained by selling one more unit at the lower price as shown by the demand curve. The rectangle *CBGF* represents a loss of receipts arising from selling *OA* units at a price *OF* instead of at *OC*, and the rectangle *ADEG* shows the increase in receipts due to the sale of one extra unit at price *OF*. The net increment in receipts, or marginal revenue, is equal to the difference between the two. Applying what we learned earlier of the relation between elasticity of demand and changes in consumers' outlay, it is seen that, if demand is inelastic in the region of the price *OC*, the rectangle *OABC* will be greater in area than *ODEF*. Consequently, *CBGF* is greater than *ADEG*, and marginal revenue is a negative quantity. Similarly, if demand is of unit elasticity, the two rectangles are of equal area and marginal revenue is consequently zero. Only in the case of elastic demand will marginal revenue have a positive value, for only then will the negative increment *CBGF* be smaller than the positive increment *ADEG*. It is to be noted, however, that marginal revenue must always be less than price, for it is equal to price minus the price reduction necessary to sell one more unit multiplied by the number of units sold.

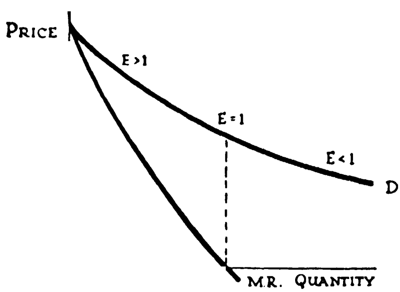


FIG. 26A.

The marginal revenue curve will therefore always lie below the demand curve, and will cross the horizontal axis when elasticity falls to unity (Fig. 26A).

For comparison, we may look at the marginal revenue curve of a firm working under conditions of perfect competition. It is then always possible to sell one more unit at market price, so that, as there is no fall in price, there is no negative item in marginal revenue, which is always equal to market price. The

individual firm's demand curve and its marginal revenue curve coincide under conditions of perfect competition, and both are horizontal straight lines (Fig. 26B).

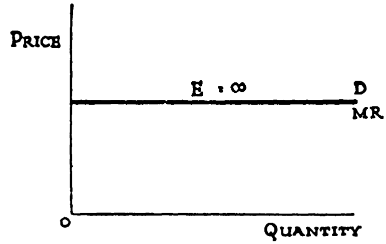


FIG. 26B.

Price, Marginal Revenue and Elasticity of Demand

It is clear from the above argument that marginal revenue is less than price, except in the case of infinite elasticity. When elasticity is equal to unity, marginal revenue is zero. It is possible to arrive at a simple formula which will give the relationship between price and marginal revenue for any value of elasticity.

Let price be p and the quantity sold q. Then for a small fall in price Δp quantity sold increases by Δq and the increment in receipts is ΔR. Marginal revenue is defined as a small change in receipts for a small change in quantity sold, and is therefore $\Delta R/\Delta q$.

Revenue at the original price is $p.q$.

Revenue after the price change is $(p - \Delta p)\ (q + \Delta q)$.

Then we have:

$$\Delta R = (p - \Delta p)\ (q + \Delta q) - p.q.$$
$$= pq + p.\Delta q - q.\Delta p - \Delta p.\Delta q - pq$$

which reduces to $p.\Delta q - q.\Delta p$, since $\Delta p.\Delta q$ can be ignored as being of the second order of small quantities.

$$\therefore \frac{\Delta R}{\Delta q} = p - \frac{q.\Delta p}{\Delta q}$$

Now elasticity of demand $= e = \dfrac{\Delta q}{q} \cdot \dfrac{p}{\Delta p}$

$$\therefore \frac{\Delta R}{\Delta q} = p - \frac{p}{e} = \frac{p(e-1)}{e}$$

Marginal revenue $=$ price $\times \dfrac{e-1}{e}$.

This equation satisfies the conditions that marginal revenue is equal to price if e is infinite and is equal to zero if e is equal to unity. For all values of elasticity greater than one and less than infinity, marginal revenue is less than price. The mar-

ginal revenue curve will therefore lie below the demand curve whenever the demand curve is downward sloping, and the extent to which it lies below it depends on the value of the fraction $\frac{e-1}{e}$.

Price Determination under Simple Monopoly

We have now the necessary data to complete our account of the pricing process under monopolistic conditions. We pictured the monopolist deciding on the size of output. He will

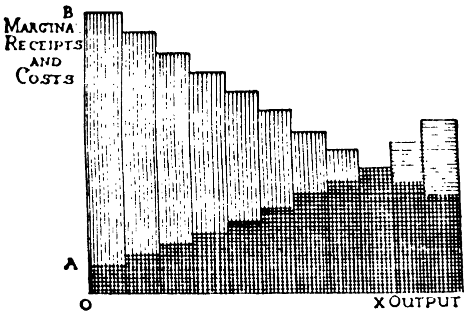


FIG. 27.

always be willing to increase output, if what he receives by way of increment in receipts is greater than the increase in his money expenses, that is, if his marginal revenue exceeds his marginal costs. This is seen in Fig. 27, where the vertically hatched rectangles represent marginal revenues and the cross-hatched portions marginal costs. So long as any net receipt item such as AB exists, it is profitable to increase output. For output OX, the increment in cost caused by adding one more unit swallows up the increment in receipts accruing from the sale of that unit and further expansion of output is not profitable. The quantity sold will therefore be determined by the intersection of the marginal cost and marginal revenue curves, for that is the point at which net receipts are a maximum. The price P which must be charged to sell this quantity OX can then be read off from the demand curve, Fig. 28.

The process is, in fact, precisely the same as in the case of perfect competition, only there marginal revenue is the same as market price and we can proceed straightway to the relation—

marginal revenue equals market price—although in doing so we are apt to lose sight of the fact that the really fundamental relation is the equality of marginal cost and marginal revenue.

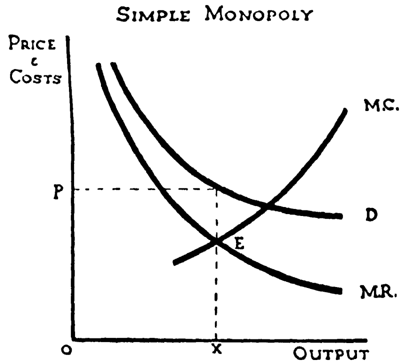


FIG. 28.

It is now necessary to enquire what the relation between price and average cost will be. In the case of simple monopoly, the monopolist is regarded as the sole supplier of a commodity; that is to say, there is nothing which is a sufficiently close substitute for his product that it must be regarded as the same commodity. The Post Office has such a monopoly of the carriage of letters in this country, the law forbidding any other body from performing the same service. Government-controlled agencies have a monopoly of telecommunications in this country and the B.B.C. has still a simple monopoly of sound broadcasting though rival television services are licensed by Government. In road transport the British Transport Commission is protected from the competition of private enterprise by statutory control of the latter. The Road and Rail Traffic Acts also serve to temper the forces of competition for the state-owned railways. Cases are less common among material goods than among services, for in the type of case cited above there is usually some legal barrier keeping out the would-be competitors. There are cases of commodities where there is not, strictly speaking, a simple monopoly, but one producer is a sufficiently large part of the industry to be able to act as if it did exist.

Further, if supplies of factor are prevented from entering an industry by legal restriction, or if an enterprise has gained possession of the whole supply of some natural resource, other firms cannot be set up to compete with the existing ones. In these circumstances, fixed cost as we used it in our discussion of pricing under competition has very little meaning. There we were able to get a market price for all factors, for even in the

cases of units of factors earning contingent rewards there was the condition that they must, in the long run, earn as much as they could expect to get elsewhere. We could include them therefore in fixed costs at a figure equal to the average earnings of similar units in other occupations. Now we have no such guide, because there are no similar units, or at least no similarly situated units.

Fixed cost can therefore include only payments to those factors which are used in fixed amounts but which the firm buys at market prices. Fixed cost as so defined, plus normal variable cost, constitutes a total cost which must be covered in the long run if the monopoly is to remain in business. Net receipts, i.e. the difference between total receipts and total cost as we have now redefined it, constitute the income of the monopolist. Since it is in general impossible to be a monopolist without possessing some factors of a specialised character, we can regard net receipts as payment to these factors which have no market price.

Completely unique factors are very rare, and even so there may be some commodity which is to some extent a substitute for the monopolised product. To this degree the net receipts of virtually every monopoly are vulnerable to competition.

Imperfect Competition

If possibilities of substitution are greater than this, as in the case of an industry composed of a small number of firms making products, which are near but not perfect substitutes for each other, it may not be possible for any one of the firms to secure an excess of receipts over total costs. The factors of production used by one firm may be substantially the same as those employed by any other, they may buy machinery from the same makers, they compete for raw materials and labour of all kinds. The products they produce differ but little so far as any physical tests can show, but in the opinion of the consumers there may be important differences. The consumer will prefer one brand to another at the same price, but other brands are sufficiently close substitutes for the existence of a price difference of more than trivial dimensions to change the direction of his purchases. There will thus be no specialised factors of production employed, since it is possible to set up a new firm

to make a product very like those other firms are already selling. Consequently if in certain firms units of any factor of production are earning more than similar units are earning elsewhere, new firms will be set up to make similar products at slightly lower prices than existing firms are charging, and any surpluses of net receipts (total receipts minus total variable costs), over the sum sufficient to pay all fixed factors their market prices

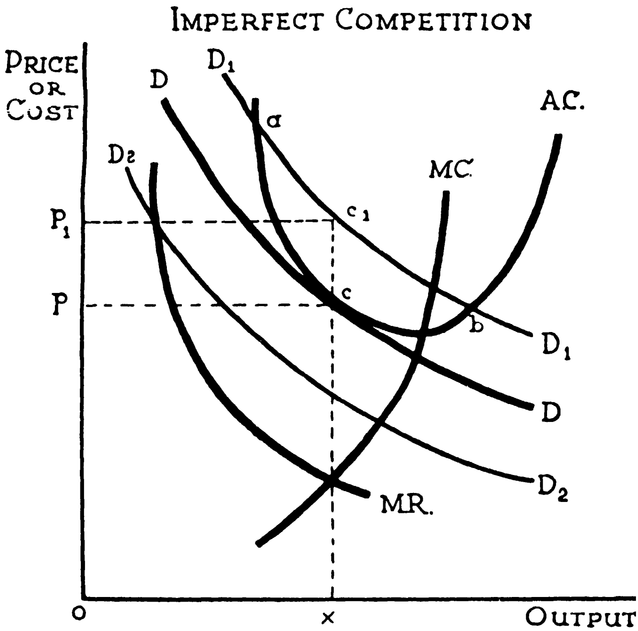


FIG. 29.

and no more, will disappear. There will be a market price even for those factors working for a conditional reward, and we are concerned with a total cost curve which is identical with that we used for a competitive market. To maximise profits or minimise losses a firm will fix output at the point where marginal cost and marginal revenue are equal, say *Ox* in Fig. 29. The figure shows three possible positions, *D*, D_1 and D_2, of the demand curve, but to avoid confusion the marginal revenue curves associated with D_1 and D_2 have been omitted. If the demand curve is in the position D_1, i.e. it cuts the average

cost curve in two points *a* and *b*, it is easy to show that profits will be maximised at some point on the demand curve between *a* and *b*. To the left of *a*, average cost is greater than price, so that no point of equilibrium is possible. At *a* price and average cost are equal and one condition of equilibrium is fulfilled, but to the right of *a* average cost falls below price, and the producer is likely to recognise this and increase output to gain a net profit on each unit sold. This margin between price and average cost increases for a time with increasing output. Similarly, if a producer finds himself at *b* on the demand curve for his product, it will not pay him to increase output, but it will pay him to decrease it. Since net profits are zero at *a* and *b* and positive in between, there must be some point where they are a maximum, but a situation in which a net profit exists cannot be permanent. Other firms will be attracted into this line of production, and will produce other products which are near substitutes for the product in question. If necessary they will charge a price low enough to attract away some of the first firm's customers. Our firm will consequently experience a decrease in demand, and its demand curve will move to the left. The argument runs on lines which closely resemble those of the analysis of price determination under conditions of perfect competition on pp. 112–118, the only difference being that, under perfect competition, a firm which tries to charge a higher price than its rivals loses all its customers immediately, while in conditions of imperfect competition, it loses some of them only; those who are least firmly attached to it depart first.

Again, just as in the former argument, it can be shown that a firm whose price is below its average total cost will tend to disappear, because in spite of the equality of marginal cost and marginal revenue, it is not able to pay normal rates of return to its fixed factors, so that a position of the demand curve such as D_2 cannot be permanent. Thus, since there can be no long-run equilibrium when the demand curve cuts the cost curve, nor when the two curves do not come into contact, there is only one alternative—the two curves must touch. At the point of osculation, price and average cost are equal, so that receipts equal total costs, but in order that the point shall be one of equilibrium, the demand curve must lie below the average

cost curve except at this point, for if it should lie above it, some output greater or less than Ox would be more profitable, as receipts would exceed total cost. If, however, the demand curve touches the average cost curve from below, as in Fig. 29, the point of contact will be one of maximum net receipts or minimum loss (both are zero), and therefore marginal cost and marginal revenue must be equal for the same size of output. The marginal curves will therefore cut at a point vertically below the point of contact of the average curves.

Because marginal cost and marginal revenue are equal, there is no tendency for the firm either to expand or contract output; because total costs are equal to total receipts, there is no tendency for existing firms to go out of business or for new ones to set up. The situation must not, however, be confused with competitive equilibrium, for price is greater than marginal cost, since it is equal to average cost which is greater than marginal cost. Average cost and marginal cost are equal only at the point where average cost is a minimum, and since the demand curve is downward sloping, it must touch the average cost curve on the downward sloping half, so that average cost is greater than marginal cost, and output is less than that which makes average cost a minimum.

This latter result is consonant with the belief of many businessmen that they could reduce costs per unit if they could sell more, but that a reduction of price would not be profitable to them. The above analysis probably provides the explanation for a large proportion of such cases. It is probably applicable to the sale of commodities sold under an advertised brand name and to retail distribution generally; indeed, in the latter case it is easy to show how the process can work without price changes.

Suppose that in a recently built garden city there are fewer grocers per thousand of population than is normally the case. The large turnovers of existing businesses may enable unusually large profits to be made, even though the prices charged are the same as elsewhere. Other grocers hear of the high level of profits and set up shops in the area. We will further suppose that customers choose their grocer solely from the point of view of distance from the shop to the home, so that each of the original grocers had an assured market consisting of all the

households for whom he was the nearest grocer. When the new grocers enter the field, they will be nearer to a certain number of households than any of the original shops, and each will establish an assured market of all the households for whom he is now the nearest grocer. Each new-comer will thus carve out a market for himself from the markets of one or more of the original shops. The clientele attached to each shop, old or new, will be smaller than formerly, and as more and more new shops are opened the assured market attached to each shrinks steadily. As the volume of business done at each shop decreases, profits decline, and the situation is eventually reached when the prospective income to be derived from setting up a shop in our garden city is no greater than that anticipated from a similar shop elsewhere. In terms of Fig. 29, the demand curve will have moved from a position such as D_1 to position D, and total receipts will just suffice to pay the factors of production now employed in the grocery trade such rewards as they could expect to earn elsewhere and no more.

Price Discrimination

So far, in each problem, whether of competition or monopoly, there has been only one price in a market for all who wished to buy, and we will now deal briefly with the circumstances in which there may be two or more prices for the same article in the same market. We saw in our study of conditions of demand that the limit above which no purchase would be made was different for different buyers in the market, so that people who would be willing to buy at much higher prices are enabled to buy more cheaply and possibly with a smaller total expenditure. If it is possible to separate different types of demand, it may be possible to charge different prices to the different types and thereby extract a larger total revenue.

Thus, railway passenger journeys can be divided into those taken in connection with earning a living and those taken for recreation. The demand for the former is likely to be very much more inelastic than for the latter, so that it may be possible to charge much higher prices for business journeys than for pleasure journeys, if it is possible to separate them. One way of separating them is to offer cheap excursion trains on Sundays when there is little demand for business journeys, and

to try, by forbidding luggage, to prevent people transferring to Sunday journeys they might take on a week-day. The demand for Sunday journeys can thus be dealt with separately, and an equilibrium between the marginal revenue obtained from them and the marginal cost of running excursion trains found. The fixed costs connected with the track, stations and rolling-stock form a large part of total costs for railways, so that if, in addition, the demand for excursion journeys is very elastic, very large price differences between excursion and ordinary tickets may exist. Similar considerations govern the offer of cheap tickets into town from suburban stations during the middle of the day. If the equipment necessary to carry the heavy traffic of the " rush " hours is not required except in the morning and evening, it will pay to induce housewives to shop at city stores instead of at suburban shops, provided the receipts from their tickets yield a margin over and above the extra expense incurred by running the trains instead of leaving them idle.

Whenever a market shows peaks of inelastic demand at certain times, then it offers opportunities for differential pricing, provided, of course, it is not possible to buy in the low-price period to use in the high-price period.

In this connection it may be noted that both gas and electricity undertakings have the same problem of a peak demand in the evening, but while electricity suppliers normally offer lower prices to capture more elastic demand at other times, gas suppliers do not, the reason being presumably that it is easier to store gas.

With material goods the possibilities of transfer of demand from the high-price market to the low-price one are much greater. An industry which, by means of a cartel and a tariff wall, has secured control of its home market and rendered the demand for its products there more inelastic by shutting out competing supplies, often finds it profitable to sell in a competitive market abroad at a lower price than at home. This practice is commonly termed " dumping at a price below cost of production," which, if it were indeed true, would mean that the concern was acting as a philanthropist. What is actually happening is that it is following the orthodox policy of equating marginal cost and marginal revenue in both markets. The

export market is competitive, so that the monopoly can always sell another unit at the market price; marginal revenue and price are thus the same. The cartel fixes its total output where marginal cost is equal to price in the export market. In the home market it sells such a quantity that marginal revenue there is equal to the marginal cost of total output. Home market price will be greater than marginal revenue, since the cartel enjoys a monopoly, and hence will also be higher than the price in the export market.

Where it is impossible to force consumers into the high-price or the low-price market, it may be possible to induce them to sort themselves out voluntarily. This is done by misleading the consumers as to the quality of the goods they are buying. The product is sold in different grades at different prices; there may be no actual difference between one grade and another, or differences in costs may be in no way proportionate to price differences. Separate demand conditions are set up for each "grade," and receipts are greater than if there were only one price in the market. Thus, in pricing theatre seats, the difference in the cost of seating people in the dress circle and in the gallery is in no way proportionate to the price difference, and social snobbery sends the right people in at the right entrances. Again, in the soap trade the difference between the prices of common household soap and expensive toilet soap is probably reflected neither in cost nor in chemical composition, but on the basis of claims made by skilful advertising, consumers are willing to sectionalise their own demand.

Recalling an earlier discussion (see p. 122), it is clear that such cases can be treated as joint products, but whereas in our examination of the pricing of joint products under conditions of perfect competition we used the horizontal demand curve for the product of the firm, we must now consider the demand curve for each joint product to be downward sloping, and hence it will have its associated marginal revenue curve. The intersection of the appropriate marginal revenue and marginal cost curves will determine the quantities of each to be produced, and the demand curves will give the prices at which these quantities will be sold in the usual manner. This form of the analysis applies only to cases where special costs attach to the production of each product. Other cases can be dealt with by

a simple extension of the method, but for the sake of those readers who are easily frightened by curves, further elaboration has been transferred to the appendix.

There is, of course, another type of differential pricing—that which occurs when a large concern with a wide market undercuts a small concern with a localised market with the object of driving it out of business. The foregoing analysis is not applicable to such a case, which falls outside the scope of this book.

Monopoly Power and Advertising

In general, some element of monopoly must be present before it pays a producer to advertise. Expenditure on advertising is incurred because it is believed that a return in increased profits will accrue to the advertiser and, if this is to come about, the advertiser's product must be distinguishable from the very similar products of his rivals. It is useless, for instance, for a sugar refiner to spend money on advertisements intended to induce people to eat more sugar. Any result would be spread over all the sugar refiners in the market, so that, if there were a large number of them, the increase in sales enjoyed by one alone might be quite small. In these circumstances the refiner might pack his sugar in a distinctive packing, give it a registered trade name, and advertise it in a manner to suggest it possesses properties altogether superior to those of the common article sold in bulk. He would then secure for himself any benefits which might accrue from his advertising, instead of dispersing them among the whole trade. Of course, if one refiner were a very large part of the sugar trade, or if the refiners all banded together, the advertisement of sugar as such might be profitable.

Effects of Advertising on Demand

The purpose of any seller who decides to advertise his product is to modify the demand curve for the product in a manner advantageous to himself, that is, in a manner which increases net receipts. In general, advertising will have the effect of enabling the seller to increase his sales without decreasing price; the quantity item against each price in the demand schedule will be increased or the demand curve will be moved to the right from position D_1 to position D_2 in Fig. 30. A new point of

equilibrium of marginal revenue and marginal cost will be found for output x_2 instead of x_1, as the marginal revenue curve will have moved with the demand curve.

At the same time, the elasticity of the curve may be altered. The seller may, by advertising, succeed in attaching consumers

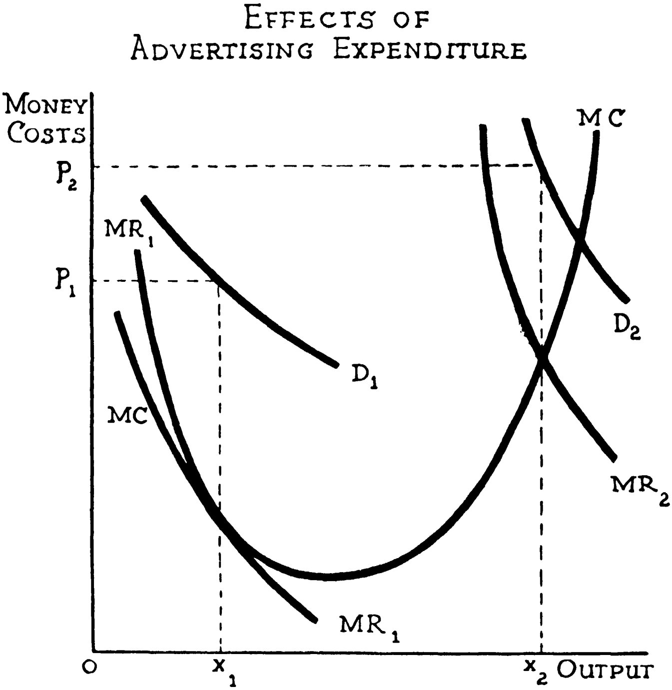


FIG. 30.

more firmly to his product, so that they will be less likely to be attracted away by price reductions or expenditures on advertising by his rivals; that is, he endeavours to make the demand for his product more inelastic for the future than it has been in the past. This same change in elasticity may enable the seller to raise his price and obtain greater net receipts for a smaller production, and consequently for a smaller outlay on resources than would have been the case if elasticity had remained unchanged.

In Fig. 30 the demand curve has both moved to the right

and become more inelastic. The marginal cost curve remains the same, since advertising expenditure does not vary in any very direct manner with output, and is best treated in this respect as a fixed cost item. The advertising campaign is worth undertaking if net receipts from selling ox_2 units at op_2 exceed net receipts from selling ox_1 at op_1 by more than the expenditure on advertisement.

Advertising Expenditure and Communal Income

In so far as advertising increases accurate knowledge of commodities available for consumption, it adds to the communal income. With it consumers get a collection of things to consume which they prefer to the collection they would get if they lacked the information they obtain from advertisements. A great deal of trade advertising falls into this category, as it is in effect a means of disseminating technical knowledge. In this way the resources which are used in the production of advertisements are making an indirect contribution to the income of the community.

With advertisements of consumers' goods, this is not invariably and perhaps not generally the case. Information contained in them is frequently misleading if not erroneous. A great deal of expenditure is also commonly incurred to induce buyers to purchase one brand of a commodity rather than another which possesses almost identical physical properties. The sellers of the second brand also advertise to prevent the first advertiser from annexing their customers. In this way, by attack and counter-attack, it is possible for advertising expenditure on both sides to mount to a considerable proportion of the price of the article without either party being very much better off in the end. In this case the whole resources of labour, capital and materials used in advertising are wasted, in that the use of them adds nothing to the income of the community.

Demand for Factors by a Monopoly

Our conclusions about the nature of the demand curve for a factor of production, when the firms demanding it are selling in perfectly competitive markets, needs some modification when some of the users are selling their products in markets in which some degree of monopoly exists. In the case of a

monopolistic producer, the range of outputs he is prepared to produce is limited to the region of the demand curve for his product in which demand is elastic, for it is in this region only that marginal revenue is positive. Since it is in the upper or left-hand end of the demand curve in general that demand is elastic, this may be taken to mean that the monopolist will not expand output beyond the point at which elasticity becomes equal to unity. He will therefore have no demand for factors of production in quantities greater than are required to make this output.

Just as in the case of perfect competition, the quantity of factor employed will be such as to equalise the marginal cost of the factor and the value of the increment in output due to the addition of the last increment of factor. The marginal cost of the factor is the same as its price, for it is purchased in a competitive market. The value of the increment of product is, however, no longer equal to marginal physical product multiplied by the price of the commodity. The firm is selling in a market which is in some degree monopolistic, and the demand curve for its product has therefore a downward slope. *The value of the marginal product will therefore be obtained by multiplying marginal physical product by marginal revenue.* Marginal revenue is always less than price, so that when a firm enjoys some degree of monopoly its demand for factors, and hence its output, is always smaller than if it were selling at the same price in a competitive market.

Appendix

The study of the behaviour of an individual firm is based on the assumption that it attempts to make net receipts a maximum. The firm's policy is the same whether it is operating in a monopolistic market or in one where perfect competition exists. Price must therefore always be such that:

$$\frac{d}{dx}[x.f(x) - \phi(x)] = 0$$

where x is the quantity sold, $f(x)$ is the price at which it can be sold, and $\phi(x)$ is the total cost of producing x units.

That is: $$\frac{d}{dx}x.f(x) - \frac{d}{dx}\phi(x) = 0 \quad . \quad . \quad (1)$$

Marginal Revenue = Marginal Cost.

In the case of a firm operating under conditions of perfect competition, there are perfect substitutes for its product, so that $e = \infty$ and the demand curve for its product is a horizontal line $p = c$, where c is a constant.

Substituting in (1):

$$\frac{d}{dx}\, cx - \frac{d}{dx}\, \phi(x) = 0$$

$$\therefore p = c = \phi'(x) \quad . \quad . \quad . \quad . \qquad (2)$$

Price = Marginal Cost.

Furthermore, it is necessary that total receipts shall equal total costs, so that all factors shall be paid their market prices.

$$\text{That is: } xp = \phi(x)$$

$$\text{or, } p = \frac{\phi(x)}{x} \quad . \quad . \quad . \qquad (3)$$

Price = Average Cost.

The conditions of equilibrium for the firm in a perfectly competitive market are therefore fully defined by (2) and (3).

Joint Products

Let A, B, C, etc., be joint products of a firm.

$F_A(x)$, $F_B(y)$, $F_C(z)$ . . . etc., their demand functions;

$\phi_A(x,y,z \ldots)$; $\phi_B(x,y,z \ldots)$; $\phi_C(x,y,z \ldots)$, etc., their cost functions.

Then the quantities of each produced are determined by:

$$\frac{d}{dx}.\, x.F_A = \frac{d}{dx}.\, \phi_A \text{ etc.}$$

Which for perfect competition reduces to

$$p_A = \frac{d}{dx}.\, \phi_A \text{ etc.}$$

Simple Monopoly

As in perfect competition, we have the condition that net receipts shall be a maximum:

$$\frac{d}{dx}\, x.f(x) - \frac{d}{dx}.\, \phi(x) = 0 \quad . \quad . \qquad (1)$$

$$\therefore f(x) + x.f'(x) - \phi'(x) = 0$$

$$\therefore \phi'(x) = f(x) + x.f'(x) = p - \frac{p}{e}$$

where $e = -\frac{f(x)}{x.f'(x)}$ = elasticity of demand.

$$\therefore p = \frac{c\phi'(x)}{e-1}$$

$$\text{Price} = \text{Marginal Cost} \times \frac{e}{e-1}.$$

Imperfect Competition

In the case of imperfect competition, there is a downward sloping demand curve for the product of an individual firm, but because the factors of production used by one firm are identical with those used by any other and available in the market, the reward earned by units of factor cannot be greater than those same units might expect to earn elsewhere. If factors are being paid more than they could earn elsewhere, new firms will be set up and the demand for the products of all existing firms will decrease as customers transfer to the new-comer. Price must therefore be equal to average cost, so that total costs are equal to total receipts.

We have, therefore:

$$x.f(x) = \phi(x);\ \text{or}\ p = \frac{\phi(x)}{x}$$

But, $p = \frac{e}{e-1}\phi'(x)$

applies as in simple monopoly.

Price and quantity can therefore be determined by the solution of the two equations:

$$p = \frac{e}{e-1}\phi'(x) \quad . \quad . \quad . \quad . \quad (1)$$

$$p = \frac{\phi(x)}{x} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Discriminating Monopoly

When separate marginal cost curves can be attributed to each of a number of joint products sold under conditions involving some degree of monopoly, use can be made of the constructions either for simple monopoly or for monopolistic competition as may be appropriate in the circumstances. The same method is applicable, even when the so-called joint

products are physically the same goods, as with electricity supplied for different purposes, or train journeys taken at different times of the day, provided always that marginal costs are separable. If a uniform product is being sold at discriminating prices and no special costs attach to the sale in any sub-market, a slight modification of this method is needed. It

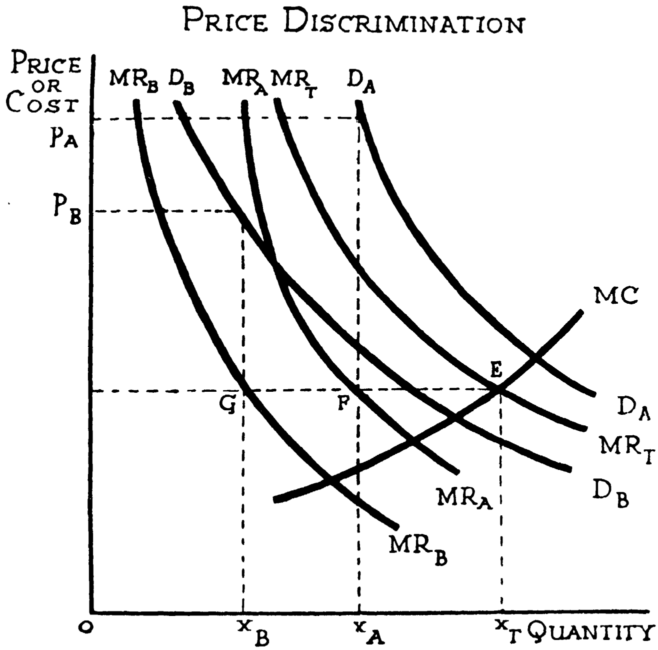


FIG. 31.

is then impossible to distinguish between the marginal cost of supply to one market or to another, but the marginal cost of the whole supply is ascertainable.

Furthermore, discrimination is possible only when the demand curves of individual purchasers are not all identical. If each individual purchaser could be dealt with separately, each would be induced to extend his purchases until marginal revenue from them is equal to marginal cost of the whole output. Marginal revenue from the purchases of each buyer would thus be the same, being equal to marginal cost. If each pur-

chaser's demand curve were the same, the price at which marginal revenue and marginal cost were equal would be the same for all, and the price which maximised net receipts for one would do so for all. If individual demand curves have different elasticities, the prices equating marginal cost and marginal revenue will differ. In general, individual treatment of consumers will not be possible, and an attempt will be made to group together individuals whose demand curves are believed to be similar.

The following analysis provides a solution for the case of two sub-markets, but the method is immediately extensible to any number.

A separate demand curve exists for each market, and each will have its marginal revenue curve. By taking the quantities sold in each market at successive values of marginal revenue and adding them, we obtain the marginal revenue for the total output of the concern. Total output will be in equilibrium when marginal revenue and marginal cost are equal. Thus, in Fig. 31 D_A and D_B are the demand curves and MR_A and MR_B are the marginal revenue curves for the markets A and B respectively. MR_T is obtained by adding the two sub-market marginal revenue curves together, and the total quantity to be produced is determined by the intersection of this curve with the marginal cost curve. If a horizontal is drawn from E, the point of intersection of the total marginal revenue curve and the marginal cost curve, to cut the sub-market marginal revenue curves in F and G, perpendiculars from these points dropped to the x axis and produced upwards to meet their respective demand curves will give the quantities sold in each of the sub-markets and the prices at which they are sold, i.e. x_A at p_A and x_B at p_B. By the definition of the total marginal revenue curve:

$$x_A + x_B = x_T$$

In this case it is not necessary that net receipts shall be a maximum in each of the separate markets in order that net receipts for the two markets combined shall be a maximum. But since the incremental cost due to any additional unit is the same, irrespective of the market in which it is sold, it is necessary that marginal receipts in each market shall be the same, for if they are not, sales will be increased exclusively in one market until the difference disappears.

A practice commonly included under the omnibus title of " dumping " can also be handled by means of this apparatus. Let us suppose that a cartel has a monopoly of a protected home market and is also selling in a competitive export market. The demand curve for its product in the home market *DH* (Fig. 32) will be downward sloping, while that in the export

DUMPING

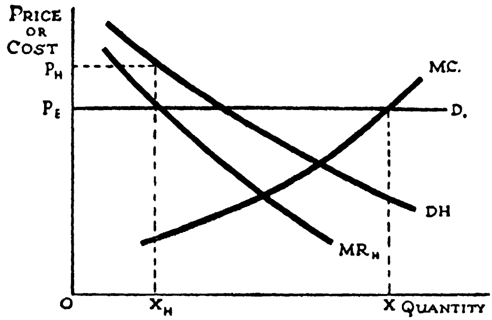


FIG. 32.

market $P_E D$ will be horizontal. As before, we will suppose that no special costs attach to selling in either market, so that marginal revenue must again be the same in either market. In the export market marginal revenue is equal to market price, so that marginal revenue in the home market must be equated to market price in the export market. Consequently, in Fig. 32 the marginal cost of the quantity Ox is equal to the export price. This will therefore be the total quantity sold, and it will be divided between the two markets, Ox_H being sold at home and the rest exported, Ox_H being that quantity the marginal revenue from which in the home market is equal to the export price.

FURTHER READING

FETTER, F. A.: *Masquerade of Monopoly.*
ROBINSON, E. A. G.: *Monopoly.*

MORE ADVANCED READING

CHAMBERLIN, E.: *Monopolistic Competition.*
ROBINSON, J.: *Imperfect Competition.*
STIGLER, G. J.: *Theory of Price*, Part 3.

Chapter XI

COST AND OUTPUT IN THE LONG RUN

OUR discussion of the relation between the size of the output of a firm and cost per unit has so far been conducted in terms of a series of outputs which are alternatives open at a moment of time, or at least within a very short period of time. We have envisaged a firm as being designed to produce an output of a certain size, but since it cannot expect the demand for its products to remain completely stable, output might be expected to fluctuate on either side of the designed output. There would be an upper limit beyond which output could not be increased at all, and there would be a lower limit below which it would be more advantageous to close down rather than to produce anything. Between these limits there would be a scale of prices and outputs forming the supply schedule for this particular productive unit.

Our hypothesis, that we are dealing with a productive unit capable of producing a limited range of outputs and able within this range to increase or decrease output at short notice, was based on the assumption that in the organisation of the firm there was at least one factor of production of which the firm was forced to use a fixed quantity, irrespective of the size of output being produced. A fixed factor of this sort must be such that the attempt to increase or decrease it involves redesigning the whole productive unit. It may be found in the plant employed, in a manufacturing process, the buildings occupied by a store, the acreage of a farm or in the individual or group of individuals managing the concern.

The existence of this fixed factor of production enables us to use the Law of Diminishing Returns to derive our relationship between cost and output, and we arrive at the idea of U-shaped average and marginal cost curves. The designed output to which we have already referred above will be that which gives

minimum average total costs. The combination of factors giving this minimum average cost depends, not only on the physical Law of Diminishing Returns which accounts for the fact that there is a minimum at all, but also on the prices of the factors employed. The Least Cost Combination of factors, which is the combination giving minimum average cost, is arrived at through the process of balancing increases in factor cost against increases in physical output, and average cost is a minimum when three conditions are satisfied:

(*a*) The average product per unit of expenditure on all factors is equal to the marginal physical return to expenditure ; or alternatively, marginal cost of output equals average cost of output.

(*b*) Equal incremental expenditures on all variable factors produce the same incremental quantity of product ; or alternatively, the ratios of the marginal physical products of the factors employed are equal to the ratios of their prices.

(*c*) The marginal physical products of all variable factors are decreasing.

There is therefore no advantage to be gained, either by varying the proportions in which factors are used, or by increasing the quantities of all variable factors proportionately. To do so will increase costs more than in proportion to output.

The set of proportions in which a given group of factors should be combined to give minimum average cost will vary with changes in the price of any factor, so that it would appear that the designed output and that which will give minimum average cost may diverge due to such changes. In general, no particular difficulties arising from this cause need be apprehended, because it is believed that average cost commonly varies but little over a wide range of output, so that the bottom of the U-shaped curve is flat.

It is on the basis of such a picture of the cost-output relationship that we have so far built up our study of the determination of prices and outputs. The scheme of analysis we have built up is, strictly speaking, applicable to conditions where time can be ignored. The Law of Diminishing Returns, on which the whole structure of deduction is based, relates to sets of alternatives which are open at a moment of time. Such limits rigidly

interpreted would be too constricting, and we have allowed rearrangements of factors which would take time, so long as the general organisation of the productive unit remained the same, or, as we have put it earlier, we have permitted changes in output which might take time, so long as the scale of output remains the same. We must now examine the changes in the cost-output relationship, when the scale of output is permitted to change.

Changes in Scale of Output Take Time

If it were worth-while in economics to aspire to the exactness of the natural sciences, we should be obliged to talk of the relationship between cost and output in terms of three-dimensional diagrams, for we have in fact three variables—cost, output and time. The changes in organisation of factors which we now have in mind take time to accomplish. Such a change is that which takes place in the development of a mining concern, when a shallow shaft served by a simple steam winch or a sloping adit with pony-hauled trucks is replaced by elaborate power-driven winding gear, and the pick and shovel and man-handled tub below ground are replaced by machine cutters and conveyors. Such a change occurred in textiles when hand looms were replaced by power-driven looms arranged in lines in larger buildings where they could be driven by belts from shafting turned by one large engine, or again when the shafting was removed and new looms installed capable of being driven at higher speed by separate electric motors supplied from a central power-house. Such changes as these take time to organise. They are introduced only if it is believed that the increase in demand is of sufficiently permanent character to justify the necessary investment.

Changing Supply Conditions

It is clear we could not represent such a growth in output by a continuous line supply curve, such as we described in Chapter VII. The technical conditions underlying supply have completely changed, and a new relationship between cost per unit and output has superseded the old one. At any given time the cost-output relationship can be represented by a U-shaped average cost curve, but at the end of a period during which

technical changes have taken place we shall be dealing with another average cost curve which will have a minimum point at a large output if scale of production is increasing, and at a smaller output if scale is decreasing. Thus, in Fig. 33, AC_1

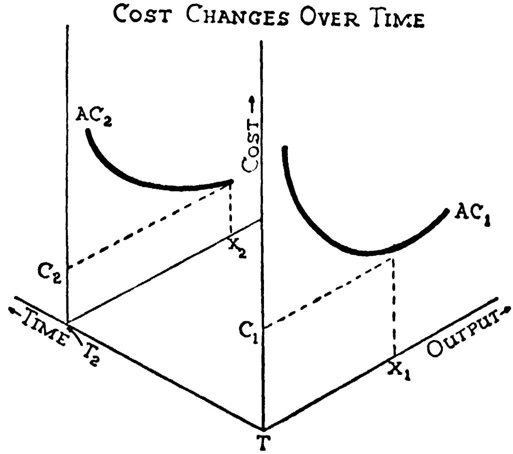


Fig. 33.

represents an average cost curve at time T_1 when quantity produced at minimum average cost C_1 is X_1, and at time T_2 the average cost curve is AC_2 and X_2 is produced at minimum average cost of C_2, which is smaller than C_1.

Varying Cost Conditions Defined

In the case illustrated diagrammatically above we have assumed that the change in technique is such that minimum average cost has fallen as the scale of production has increased. It is quite possible that minimum average cost may increase, decrease or remain constant, and according to the condition which rules, we say that we are concerned with increasing, decreasing or constant cost conditions. It seems best to use the term cost conditions, rather than to talk of increasing, decreasing or constant cost; or of decreasing, increasing or constant returns, as once was customary. Older writers make no distinction in analysis between changes in cost per unit due to the varying proportions in which factors are used, scale remaining constant, and changes in cost per unit due to changes

in scale. Many adopted the unfortunate practice of talking about decreasing costs when referring to the downward sloping part of the average cost curve, and of increasing costs when talking of the rising portion. The terms increasing and decreasing cost conditions indicate the essentially discontinuous character of the change in costs.

Indivisibility of Factors

Another important feature of the cost-output relationship is that it may be possible to increase the quantity of one or more factors employed only by adding doses of some considerable size. Thus, suppose it were possible to obtain land for farming only in blocks of 500 acres. There would be some size of output from a 500-acre farm which would give minimum average cost, and for larger outputs cost per unit would rise, but for some range of output a one-block farm would still be the most economical production unit. As operations further expanded, an output would be reached which could be produced equally efficiently by a one-block farm or a two-block farm, but, for larger outputs still, a two-block farm would be the more economical method of production. This process is illustrated in Fig. 34A. Average cost will be at the same minimum level when factors are combined in the same set of proportions. The discontinuity in this case is due solely to the indivisibility of a factor of production. The same phenomenon might arise in a manufacturing process where " lines " of plant have to be used, that is, where a series of machines work on the output of a single machine. A single line of plant may be a perfectly efficient productive unit, but no increase in capacity short of doubling by adding a second line can be made.

If this indivisibility of factors could be removed, then it would be possible to increase the use of all factors by any amount, however small, and to maintain the most efficient combination of factors. So long as this set of proportions was maintained, average cost would remain constant. It is not, however, likely that cost could remain constant indefinitely, irrespective of size of output. Sooner or later the supply of some factor of production would become restricted, and would be available only in quantity insufficient to keep the proportion constant, or the price per unit of the factor would begin to rise,

or the quality of it to fall off. Any one of these would bring the state of constant cost to an end, and increasing cost would take its place.

It might, however, be that as output is increased certain economies become possible. It might be that as output is

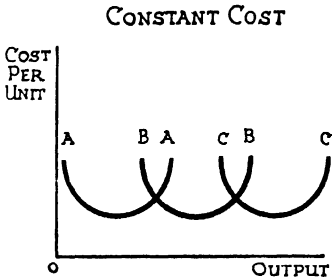


FIG. 34A.

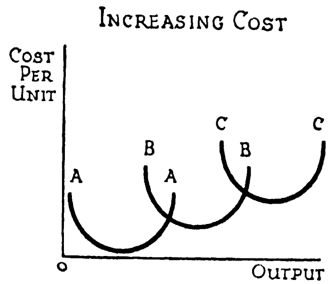


FIG. 34B.

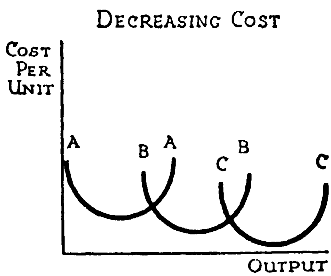


FIG. 34C.

increased use can be made of by-products. Then, although quantities of factor used per unit of physical output and all factor prices remain the same, economic output is greater; for certain things, which were formerly valueless, now have a value. Some of the costs can now be attributed to the by-products which were formerly waste, and costs per unit of the main product are correspondingly reduced.

It is to be noted that what may be a cheap method of producing a large output may be a dear method of producing a small output. Thus, the average cost of producing certain outputs as shown by the curve *CC* in Fig. 34C is greater than the average cost of the same quantity according to either *AA* or *BB*. The reason for this is clear if we consider how expen-

sive it would be to produce an output of 10 cars a week on a plant designed to produce 1,000. This feature is commonly found in cases of decreasing cost conditions.

Increasing Cost Conditions

We have already seen that a condition of increasing cost can arise when the supply of at least one factor is restricted, and an increased demand for it on the part of users as a whole will send up its price. An extreme example of increasing cost conditions is given by the case on which Ricardo founded his theory of rent. The quantity of agricultural land is taken as fixed and all of it is occupied. Then, if population grows, so that the demand for food-stuffs increases, more can be provided only by extracting more from the land already in cultivation. Competition for the use of land will drive up rents, and any net receipts accruing due to the rising prices of food-stuffs will be absorbed by the rise in the rent of land.

If we look at the problem from the point of view of the individual farmer,[1] we shall see that the price of food-stuffs will follow the marginal cost curve, and since there is competition no excess of receipts over costs can exist for long. Now, when we were discussing competitive equilibrium of the firm in Chapter IX, p. 120, we saw that if firms making a particular product found themselves in a position where receipts were more than sufficient to pay all factors of production their market prices, other firms would set up; the supply on the market would increase and price fall, until receipts of each firm just covered market-price payments to all factors once more. In the case under consideration, an increase in productive capacity is impossible, as there is no unoccupied land. Land is the scarcest factor, and competition for land will drive up its price until the higher level of receipts is just sufficient to pay factors their market prices. The net receipts will have been absorbed as a rise in rent.

Price will therefore be equal to average cost as well as marginal cost, which means that the average cost curve will rise up the marginal cost curve. Thus in Fig. 35, as, owing to increased demand, price rises from P_1 to P_2, the average cost curve rises from the position AC_1 to position AC_2.

[1] Assuming that the number of farmers and the sizes of farms both remain constant.

Increasing cost conditions will also exist if, instead of a factor whose supply is fixed, we have one whose supply can be increased, given time, but at an increasing cost per unit. Then, as the demand for the commodity increases and the demand for the factor increases, firms will have to pay a higher price for the factor, irrespective of the quantity of it they use. Therefore, as the demand for this factor increases, the marginal and average cost scales of all producers will be uniformly increased and the cost curves will rise parallel to the cost axis. Strictly, we should draw such a diagram in three dimensions, but we can manage with two, provided we put a time label on each pair of cost curves.

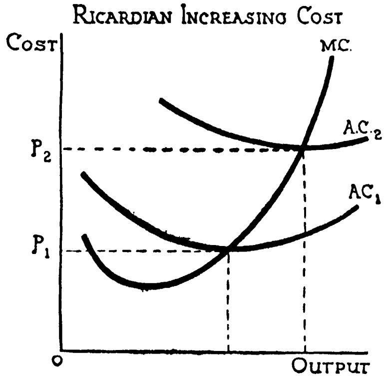


Fig. 35.

Another case which might be included in the same category is that in which a natural resource is being exhausted, but where successively inferior or less accessible supplies are being tapped. It is probably better to look at it from the point of view of a reduction in the technical efficiency of factors of production. As inferior resources are exploited, a new and lower diminishing returns curve must be used to describe the situation. Such circumstances are sometimes to be found in mining. When the more accessible seams have been worked out, the deeper ones may have to be opened, and this involves the greater cost of the deeper shaft and greater expense for haulage, ventilation and pumping of water. An explanation of the rent of land was once attempted on the basis of a similar but wholly inaccurate account of a process in which the most fertile land was cultivated first and the next grade next and so on.

Decreasing Cost Conditions

When dealing with the causes of decreasing cost conditions, it is necessary to distinguish between economies arising from a reorganisation of the factors of production inside a particular

firm and economies which the firm enjoys although it is not responsible for initiating them. It was once customary to classify economies as internal or external, from the point of view of the industry; but it is now clear that this distinction is of no analytical significance, particularly since most economists have abandoned the attempt to find a rigid definition of an industry. The distinction between economies which are internal or external from the point of view of the firm is important, and it is with reference to the firm that these terms must be interpreted in what follows. We will consider external economies briefly before giving internal economies the somewhat more detailed treatment they demand.

External Economies

First there are the increases in the technical efficiency of factors generally, due to the increased stock of knowledge and skill. Much of the progress of the past century and a half centres on technical innovations which have enabled a higher percentage of the energy stored in fuels to be turned into useful work. As a result, costs have been reduced wherever motive power is used. Certain economies are dependent on the growth of the industries to which they accrue.

Thus, when the demand for a raw material on the part of an industry reaches sufficient proportions, it may pay the industry supplying it to adapt the material more closely to the particular purpose for which it is being used. For example, when the motor industry was smaller, it used mild-steel plates, such as were being produced for a variety of purposes, but when its demand increased, it became possible for the steel industry to produce a particularly fine-grained plate on which the requisite surface could be obtained with less enamel.

Again, as an industry grows it may set up institutions which benefit all firms engaged in it. The establishment of an organised exchange, either for its product or for a raw material, may reduce the risks and interest cost of carrying stocks; research associations may be established to deal with the common problems of all members; employers' associations and trade journals disseminate information.

Further, as the industry grows, a labour force, skilled in its processes, becomes attached to it. Young labour is brought

up in the atmosphere of the industry, and technical knowledge is handed on from one generation to another more readily. The atmosphere of the industry may be such that it is highly undesirable socially that young people should be brought up in it, and the same factor tends to direct a steady stream of young labour towards an industry which is declining and can never give all of them regular employment, but at the moment we are concerned only with the fact that a certain element of economy in production is to be found in this circumstance. There is also the important feature that when a substantial proportion of a population is engaged in an industry, technical training for that industry is likely to be subsidised from public funds.

Such considerations play an important part in determining the historical trend of costs in an industry, but for reasons we will develop later, they do not lend themselves to precise analytical treatment.

It is of course possible that there might be external diseconomies, though, in general, circumstances would approximate to the case of restricted availability of one or more factors with which we have already dealt.

Internal Economies of Scale

Just as it is possible to regard increasing cost conditions as being due to the greater relative scarcity of one factor of production, so those causes of decreasing cost conditions which are internal to the firm can be subsumed under the general heading of division of labour, and two types of division can be identified. In the type described by Adam Smith in his immortal account of a pin factory, progress takes the form of further and further subdivision of human tasks. The peak of achievement in this direction may have been reached in Ford's division of the process of making the T model into 7,882 tasks. The other type takes the form of a reallocation between different types of firms and different industries of the whole chain of processes which result in a given commodity. It has been described as the Integration of Processes. We will describe it as the modern type of division of labour, although it is probably as old as the other, because it has become so prominent a feature of industrial reorganisation in the past half-century.

Division of Labour as Subdivision of Tasks

In contemporary mass-production methods, the task of an individual worker is reduced to a few simple operations, or even to a single operation, repeated in identical circumstances. This contrasts with the work of the skilled craftsman, who has to perform a wide variety of operations. The materials the craftsman works on are not entirely uniform, as are those of the mass-production worker, and each operation may call for the exercise of skill and judgment. Consequently, the mass-production worker may reach maximum dexterity in a matter of weeks, while years must pass before the craftsman can be considered fully skilled. In addition to facilitating the acquisition of a higher degree of skill, this reduction in the number of tasks a worker has to perform eliminates the waste of time due to the warming-up period, which occurs each time a change of job takes place. It is, of course, possible for a task to be made so monotonous that a change of work for part of the day is advantageous both from the point of view of the worker and of the size of output.

Again, division of labour enables use to be made of differences in individual ability by allowing workers to specialise in the field where their skill can best be utilised. This is not the same thing as saying that each individual does that work at which his skill is greatest. The editor of a newspaper might, at the same time, be its best linotype operator, but it would be foolish to transfer him from the editorial room if he could make a larger contribution to the output of the concern there. Ford carried this aspect of the Detroit organisation so far that tasks were found which blind workers could perform better than sighted, because the fatigue arising from working by touch was less than that of working by sight.

There is generally considerable scope for these economies on the managerial side of a business. When one man is concerned with the technical problems of production, the supervision of labour, the selling of the product and financial control, it is improbable that he will have the type of ability which is ideally suited to every one of these tasks, or that he will have time to give any one aspect the study it requires. Time and effort will be wasted in transferring attention from one type of problem to another.

Specialisation of workers can be accompanied by specialisation of equipment. If a worker is to be employed continuously at a particular task, the equipment he uses can be designed to give the greatest efficiency in that task, irrespective of its adaptability for other tasks. If specifically designed equipment cannot be fully employed, it may be necessary to use tools which, although technically less efficient, are more adaptable, because of the capital costs of equipment which is used only intermittently.

As the scale of operations grows, it also becomes profitable to specialise some labour on such tasks as research to improve methods of production or to devise new products; on training new labour instead of leaving new-comers to pick up their duties as best they can; on elaborate costing systems and financial control to secure the efficiency of the organisation of the factors employed.

It is worth noting in passing that many of these technical innovations on the administrative side are not a direct source of any superiority of the large-scale business over the small. They become necessary when, *for some other reason*, a business grows and creates administrative problems which do not exist in a small business.

Division of Labour as Integration of Processes

There is also a tendency to reverse the procedure of subdividing a process into many individual tasks and to have a number of operations performed simultaneously by a machine. We thus get division of labour between firms or industries instead of inside a firm. If the machine-tool industry produces a press which will make half a car body in one operation, some of the work of making cars is virtually transferred from the automobile industry to the machine-tool industry.

This problem, although it cuts across the boundaries of individual firms, is essentially a problem of the firm in a way the problems classed earlier as external were not. The scope of the operations carried out by a firm must now be considered, and we shall observe that the range is sometimes increased and sometimes decreased. At the present time in retail distribution there appears to be in progress a procedure termed " cutting out the middle man." It is often claimed that prices can be

lowered if the manufacturer sells direct to the retailer or to the public through his own shops, because "intermediate profits are eliminated." Now, no one is in the position of receiving a "profit" without performing some service for it. It may be that some people are able to make charges for their services which others consider exorbitant, but the service is there just the same, and if that link in the distributive chain is to be eliminated, someone else must perform it instead.

An example may help us; when much clothing was made at home or by small tailoring and dressmaking businesses, there was scope for the merchant whose principal asset was his knowledge of his customers' requirements on the one hand and the products of manufacturers on the other. Now, when customers are being visited, the cost of selling a large quantity of goods is little different from the cost of selling a small quantity. Thus, if a merchant sold small quantities of the goods of a number of manufacturers to a shopkeeper, he might be able to do so for a total cost no greater than any one of the manufacturers would have incurred in selling his quota. Then the lower income strata of the community, who had not bothered about fashions, began to demand cheap fashionable clothing of a type sufficiently standardised to be made on mass-production lines. The number of buyers of piece goods was reduced; the large-scale makers-up of clothing could buy in quantities large enough to enter into contracts with manufacturers; they employed skilled buyers who did not require the technical assistance the merchant very often supplied to his customers. Thus, the merchant in piece goods is being driven out of business, although he still manages to survive in certain branches of the textile trade, where the special services he can render are required. Elsewhere a body of more narrowly specialised sellers attached to manufacturing firms is doing his work.

Again, in the iron and steel industry, formerly the coke might be made by one undertaking, the iron smelted by another, and the iron remelted and turned into steel by another. By the development of "Warmewirtschaft" during this century, the processes are carried out by the same firm. The gas from the coke-ovens can be used to raise power and to heat furnaces, the spent gases from the blast furnaces can be

similarly utilised, while the iron is converted into steel without being permitted to solidify, and the steel may be rolled before cooling in mills driven by power raised by gas from the coke-ovens. The process may be taken one step forward by siting the works alongside the automobile works or the shipyard which is to use the steel, or backward by siting it at the pithead. Clearly such integrations are not likely to be made where there are risks of a stoppage of one link bringing the whole structure to a standstill. Thus, a steel-works-automobile-plant integration is likely to be more successful than a steel-works-shipyard-merger, because the demand for cars is much steadier than that for new ships. The steps which have led to this change in structure have been much more complex than in the case of the textile industry: technical innovations in blast-furnace practice and in steel metallurgy have combined with changes in the nature of demand to modify the organisation of iron and steel firms.

Whatever the technical economies of heat-saving methods may be, they cannot be reaped unless a plant can be set to work to meet a large demand for a narrow range of products. If a variety of types of steel is required, it pays to have smaller furnaces. The heat-saving process is designed for continuous working, and a change in the type of section to be rolled means a stoppage of the rolling mill while the rolls are being changed. This means that the whole plant has to slow down to prevent unusable stocks of ingot accumulating while the mills cannot take them. The extent to which it pays to introduce continuous working, therefore, depends on the volume and regularity of a demand.

Each storey in an integrated structure becomes specialised to meet the requirements of the one above it instead of having to diversify itself to meet the requirements of a number of customers, and this comes about, not because there are any intermediate profits to be eliminated, but because changes in demand or in technique have made a new combination of factors of production more efficient than the old.

The process is not always one of integration, however. It may equally well be one of disintegration. It may be that some technical innovation requires a larger unit of equipment than any one firm can keep in continuous employment. If, in

addition, its capital cost is very high, the particular process will tend to split away from the firms of which it was a part, and specialised firms, performing it alone, will be set up. This has happened in the textile finishing trades, and in the supply of certain accessories and parts in the automobile industry.

It should be mentioned that a third type of division of labour, division on a geographical basis, is commonly recognised. A study of this type resolves itself into an examination of the consequences of the fact that certain factors of production are to be found only in certain places and that labour is not perfectly mobile geographically. Attention has been concentrated on international division of labour, and in consequence special prominence is given to the problems arising from the existence of national currency systems. This type does not give rise to any special problems affecting the present argument.

Limitation of Division of Labour

There is not unlimited scope for division of labour, and it is just as true as in Smith's day that it is limited by the width of the market. If a process is broken up into 200 continuous operations, at least 200 workers must be employed, and they will by hypothesis produce a greater output per head than if employed in any process which is less divided. There is the certainty of reducing costs if a process divided into 150 operations is further divided into 200 operations, but conditions of demand determine whether it is possible to sell the increased output at the lower reserve price. If the reduction in price, necessary to make demand extend sufficiently to absorb the increased output, is greater than the reduction in cost which the economies bring, then this measure of division, though technically possible, is economically impossible, and it does not take place.

Limitation of the extent of division of labour by the conditions of demand is likely to be particularly important under conditions of imperfect competition, and such conditions probably apply to a very large field of goods, perhaps to all consumption goods and to some producers' goods as well. Furthermore, mass-production methods produce a standardised

product; if variety is required, a lesser degree of division of labour is generally found to be cheaper. Thus, not only will the ABC Motor Company have its field of potential customers limited to those car buyers who prefer its product to the very similar cars sold at very similar prices by rival manufacturers, but also it finds it can sell more cars if it makes a number of slightly different models than if it offers one only, and it must select its production methods accordingly. In this connection it is worth recalling that the T-model Ford car represented a degree of standardisation which not even the firm that introduced it has cared to repeat.

Amalgamation and Rationalisation

Because of the limit set on the growth of a firm by its downward sloping demand curve, attempts have frequently been made to find a way to escape it by amalgamating two or more firms selling products which are close rivals. In this process of horizontal integration, as it is termed, it may be possible to reduce the number of variant brands produced, and thus to increase the scale of production of those which are retained. This is possible, but on appealing to experience one is forcibly struck by the numerous instances, ranging from soap to motor-cars, where it has not proved practicable to eliminate brands.

In recent years it has frequently been urged that the Government should coerce producers in certain industries to integrate. The motives for advocating this policy of so-called rationalisation have varied, but whatever the purpose it is sought to achieve, the argument always begins with the assumption that horizontal integrations will eliminate " the wastes of competition " and enable costs to be reduced. There is of course the immediate question why, if economies of amalgamation exist, the members of the industry need coercing into amalgamation? It has been customary to explain this in terms of the narrow-minded individualism of the British businessman who is supposed to prefer to be sole master of a small business struggling to remain solvent to being an official receiving orders in a larger, prosperous concern.

This reply still begs the question of the prosperity of the rationalised concern. It is possible that each producer is

working under conditions of decreasing marginal cost, but marginal revenue is falling more rapidly than marginal cost and so he cannot expand output. If a number of firms were amalgamated and the marginal cost curve for the amalgamation was the same as that of any one of them, it is possible that the resultant shifting of the marginal revenue curve to the right would enable a larger total output to be produced and sold at a lower price. Such a situation is theoretically possible, but it is doubtful if a point of equilibrium, where marginal cost is falling, is very probable, for it is generally believed that marginal cost curves are shaped like a flat-bottomed U with almost vertical sides. To say that a particular condition of costs can exist is a different matter from saying it does exist, and even where it does exist there is no guarantee it will lead to a reduction in price.

The matter can be settled only by an appeal to facts, and a study of the history of the trust movement does not support the claim. Writers on trust problems appear unanimous in the belief that these combinations have not produced the economies which their promoters said they would achieve, and that such prophecies were often made with intent to distract attention from the monopolistic situation which was being created. The general conclusion of a study of American trusts is that combinations do not bring economies which an individual firm cannot achieve through its own growth. The success of combinations appears to be closely related to their ability to create a monopolistic situation which can be exploited.

The Nature of Long-run Cost Movements Reconsidered

It became clear at an early stage in this discussion that we cannot treat the long-run relation between cost and output for a firm in the same way that we can deal with the short-run relationship. The short-run relationship can be represented by a single curve, but in the long run we are dealing with changes in the shape and position of the short-run curve. So, although a curve purporting to show the relationship between cost and output in the long run is to be found in older textbooks under the label of Long-run Cost Curve, it is not legitimate so to describe it. Such a curve can be nothing more than an historial curve: a record of what actually happened over a

period of time, which gives no clue as to what might have happened had circumstances been a little different.

If we are to use a supply curve in conjunction with a demand curve, which is the only reason for constructing it, then it too must represent a series of alternatives, and this a long-run curve can never do. If, as a result of increasing output, some technical discovery is made or some marketing institution set up, the resultant economies continue to be enjoyed, even though output may fall below the level at which the economy came into existence. So the long-run cost curve is not reversible. A step in output may involve different cost changes according to the direction in which it is taken.

Further, when the long run is considered, the relation between cost and output is not unique. Not merely is the size of the output change relevant to the change in costs, but so also is the length of the time period in which the change is to take place. A change in output will not be met in the same way if it takes place over six months as if it took place over six years. Not only does the longer time period give more opportunity for innovation, but the methods used to increase output rapidly may raise costs permanently for the future. This is the case particularly where natural resources are permanently impaired; in mining, where hasty development methods, which concentrate on the richest seams, may make other mineral more expensive or even impossible to extract; in agriculture where "soil-mining" may destroy natural fertility and lead to erosion and desiccation. The same principle applies to a rapid expansion of manufacturing output leading to faulty construction and layout which permanently raises the level of costs.

Such considerations make it impossible to draw a long-run cost curve, and all that has been said in this chapter on the long-run influences affecting costs must be taken to relate to shifts in the position of the short-run cost curve and to changes in its shape.

Income and Cost Changes

Instead of conducting this discussion in terms of costs, we might have done so in terms of the inverse concept, productivity. Our reason for doing so is really a practical one: we

need this data of price-cost-output relationships for our study of price movements, but when we enquire into the effect of technical changes on the income of the community, we find the productivity aspect more useful. The effect of technical changes is to modify the shape and position of the productivity curves, which represent the laws of returns for the particular pairs of factors and products to which they relate.

If a technical change is such that we should term it an economy, the law of return for a factor and product to which it is applicable will be modified, so that the maximum level of output per unit of factor is greater with the new technique than without it. It is not necessary that output per unit of factor shall be greater for all quantities of factor; the new technique will be judged by comparing results when it is working at its most efficient level with other methods working at their most efficient levels.

Neither is it necessary that maximum output per unit shall require a larger quantity of the variable factor than the older method. The effect of the principle of division of labour tends to make cases of economies of large scale more common than cases of economies of small scale, but the latter are by no means non-existent. Technical changes may also bring dis-economies and output per unit of factor employed may decrease.

The effect then of technical changes of all sorts will be to modify the possibilities of obtaining output or income from a given expenditure of resources. If, as is the more common case, we are dealing with economies, the technical changes make it possible for the community to enjoy more goods and services for a given expenditure of human effort. The marginal productivity of human labour is raised. It does not necessarily follow that the income of the community will in fact increase. The technical change makes such an increase possible if the community wishes to have it. The question of the choice which is made must be left for later discussion.

But although income of the community as a whole will increase unless the members of the community prefer more leisure to an increase in income, it is quite possible that some members of the community may have a decrease in income forced upon them. Although the general effect of the technical change may be to increase the marginal productivity of human

labour as a whole, the demand for labour after it has taken place may be for a new kind of labour. The marginal productivity of the labour formerly employed may be reduced, it may indeed fall to zero, so that this body of labour suffers a reduction in income.

Further Reading

CLARK, J. M.: *Economics of Overhead Costs.*
MARSHALL, A.: *Principles of Economics*, Book 5.

More Advanced Reading

LEWIS, W. A.: *Overhead Costs.*

Chapter XII

LAND AND RENT

In most countries of the world land is privately owned, and can be bought and sold or inherited in the same way as other forms of property. This has not always been so, and the time when some form of communal ownership existed is still sufficiently recent to cause the justification of the change to private ownership to be called in question. In England, where the practice of letting land for rent was established earlier than in most European countries, interest in the social problems arising from the system of land tenure led economists to devote a very considerable amount of attention to the theoretical aspects of rent. Possibly it was preoccupation with the social problems of land tenure which made them consider that the payment for land required a different explanation from that of the payment for any other good or service.

Land was said to be peculiar in two respects: first, that the total supply of it was fixed, and second, that it was not of uniform quality. Quite apart from the first aspect of land requiring any special treatment, it means from our point of view that we need not go beyond the analysis of the allocation of a fixed quantity of factor among its various uses in the case of land, while we have to take a further step and investigate the causes of changes in the total supply in the case of the other factors. Furthermore, land is not peculiar in this respect, for if we are considering changes in output of commodities over periods of time sufficiently short, we shall find that the supply of all factors is fixed.

It is equally true of labour as of land that every unit is not a perfect substitute for every other unit. The logical way to deal with this is to group units of land into sub-factor groups, such that each contains units which are of uniform quality.

Classical Theory of Rent

Rent was defined as the payment for the original unimproved qualities of land. It would thus include natural soil fertility, mineral deposits, climatic conditions, its position and what the geographers would term its physical features. Any improvements to the land were termed capital, and remuneration for them was interest.

In the original form in which Ricardo stated it, the logic of the classical theory of rent was unassailable, though one might be permitted to doubt its usefulness. Ricardo considered land as agricultural land of uniform quality, and he combined all agricultural produce into the one category, " corn." Then he stated that rent was a surplus left over after all the other factors had been paid their market prices. If we examine this proposition with the help of marginal productivity analysis, we arrive at the same result. Each factor, other than land, will be paid the value of its marginal product, which will be equal to the value of its marginal product in other occupations. The supply of land cannot be varied, *hence it is impossible to talk about its marginal product.* If the supply of land can be neither increased nor decreased, we cannot attribute any increment of product to increases or decreases in its quantity. In mathematical language, it is impossible to differentiate with respect to a constant. Consequently, the rent of land is determined by the price of corn; it is what is left over out of total receipts after all other factors have been paid at the rate of the value of their marginal products. *Within the assumptions of the argument,* it is correct to say that rent does not enter into, i.e. does not determine, price.

When we come to consider the various uses for land, the situation is changed. Marshall discusses the effect on the price of oats of an increase in the demand for land for other uses. The rise in the rental value of land will make it worth-while to force larger crops out of the remaining wheat-land by the use of larger quantities of other factors. The " marginal expenses " of oats, and hence their price, will therefore rise. This is a statement with which we should agree, until we discover that " marginal expenses " does not mean the same thing as marginal cost. Reading marginal cost into the statement, we are prepared to

accept it, and we recognise that marginal cost will include payments to land, for, by the conditions of the problem, land is one of the variable factors. Where our marginal cost is concerned simply with changes in total costs, Marshall's "marginal expenses" was related to changes in the total of payments for all efforts and sacrifices the productive process entailed, thus excluding payments for land. His analysis was based on the subjective view of real costs which we have already rejected,[1] but once payments for the use of land are regarded as being identical in character with payments to any other factor of production, the Marshallian method becomes identical with the one we have used.

Rent and Marginal Productivity of Land

The problem of rent thus comes to resemble the problem of wages. Land, like labour, is a heterogeneous factor, and the alternative uses open to a given acre of land are not the same as those for an acre somewhere else, just as the fields of occupation open to two men may be different. There is one difference which has important social implications, although it is of no particular significance analytically: an agricultural labourer is unlikely to become a barrister without some considerable effort on his own part, but a wheat-field may become an eligible building site without any effort on the part of its owner.

The rent offered for a piece of land will be the value of its marginal product in the use to which the would-be tenant wishes to put it. Bids based on marginal productivities of various amounts of land will make up the demand schedule for land of a certain type. In general, we may suppose that the higher the bids for leases, the less land will be kept for the personal enjoyment of the owners. The private park may remain inviolate for some time as the growth of a town gradually engulfs it, but as population in the surrounding district grows more dense, its potential marginal product in other uses increases, its desirability as a private pleasaunce declines, and it becomes a housing or trading estate. The inelasticity of demand of landowners for land for their own enjoyment is responsible for the preservation of many open spaces in large

[1] See Chapter VIII, appendix.

towns. By the time the demand of the landlord had become more elastic, a public demand for parks had grown up to take its place.

The encroachment of a town into what was formerly a rural area changes the demands for land both through the growth of land for building purposes and through an increase in the demand for agricultural products. The demand for milk and fresh vegetables in the town makes these products more profitable than cereal crops, and rents rise in consequence as land is turned over to market gardens and dairy farms—forms of enterprise in which it is advantageous to be sited close to the consumer. But how, then, is one to account for the fact that in such areas it is possible to rent land more cheaply for agricultural purposes than for building sites? Land for agricultural purposes is normally let for short periods of time, while in the case of building sites a ninety-nine-year lease is common. In that time the value of the land may have risen considerably, and the owner will demand some compensation for losing control of his property for a long period of time. The price at which he sells or lets on long lease will discount the possibilities of future rises in value as they appear at the time of sale. Mistakes will be made of course; the future value may be over- or under-estimated. There are, on the coasts of this island, a number of villages which enthusiastic speculators in real estate once believed would become popular holiday resorts. On the other hand, the " unearned increment " in land values due to urban development has swollen costs of public acquisition of land to such a degree that legislation has recently been enacted to secure public ownership of any future increment which accrues when land is utilised for a new purpose.

The Differential Aspect of Rent

An alternative explanation of rent begins with the idea that there is land whose natural qualities are so poor as not to be worth paying for. Such marginal land, as it is termed, would be worth cultivating only if the use of its natural qualities could be obtained for nothing. This does not mean, of course, that no payment is actually made for the use of this land. The land is " rentless " in the strict classical meaning of the term rent—the payment for the original qualities of land. Capital may

have been invested in it by the landlord and an interest payment on this will be demanded. Land of slightly better quality than this land on the rentless margin will command a slightly higher payment, the difference being equal to the difference in the value of the product which a given amount of labour and capital could produce on this land and what it could produce on the poorer land. So we proceed, considering land of successively superior qualities. For each, tenants will be willing to offer a " rent " equal to the difference between what their resources of labour and capital can produce on it and on land on the rentless margin.

This explanation of rent is not strictly accurate, but even if it were, it would be open to the objection that it introduces a second meaning of the word " marginal," and it is highly desirable that scientific terms should have unique meanings. Hitherto it has always indicated a rate of change in a quantity with respect to another quantity, but here it indicates members of a series of units of factor arranged in descending order of quality which are only worth using if they can be had for nothing. We could overcome this objection by inventing another term, but there is no purpose in multiplying hypotheses in this fashion, when the same result can be obtained more simply by a more general hypothesis.

The so-called " differential aspect of rent " is merely a clumsy way of explaining differences in the marginal productivities of different kinds of land. If we arrange units of land into groups in which one unit is so like any other unit that a would-be tenant is indifferent as to which parcel he takes from any one group, then rent will in each case be equal to the value of the marginal product. This result has a superficial resemblance to that obtained by the differential method, but it is not the same. Rent, as a differential payment, was equal to the difference in the products obtained by the same amounts of labour and capital on the land in question and on marginal land. The quantity of labour and capital per acre will not be the same in the two cases. On more fertile land, the marginal productivity of a given quantity of capital or labour will be greater, so that more of it will be used per acre than where fertility is less, since the marginal productivities of units of a factor must be the same in alternative occupations.

We may therefore conclude that there is nothing in the nature of the payment for the use of land which necessitates special analytical treatment. Neither is the distinction between "economic rent," or the payment for the original unimproved qualities of land, and the actual payment made by the tenant, necessary. The distinction is likely to cause confusion when it is found that original and unimproved qualities can in fact change; that the position value of a site improves as a growing town approaches it. The actual payment made for a piece of land normally includes some interest on capital, which has been invested either in improvements to the land, or in buildings erected upon it. As will be seen from the subsequent argument, there is in the general case no need to distinguish between the two.

Rent as the Payment to a Factor in Fixed Supply

We have already seen that if we think of the whole supply of land being used for one purpose, the fact that the supply is fixed renders marginal productivity analysis inapplicable to the question of its remuneration. The same is true of any factor of production if the supply of it in a particular occupation is fixed.

Suppose that an hotel is built in a situation which the promoters of the scheme believe will be popular with holiday-makers. When the hotel is opened, it does not prove as popular as the promoters expected. If the hotel is to remain open, the labour employed will have to be paid as much as it could expect to earn elsewhere, and the quantity employed will be adjusted so that the value of the marginal product of labour is equal to the wage paid. Market prices will have to be paid for all materials purchased; indeed, all factors which will not work unless they get their market rate will get this rate if the hotel is to be kept open. But, since the promoters have over-estimated the demand for accommodation, it is improbable that the balance left over out of total receipts, when all these payments have been met, will be sufficient to pay a reward to capital equal to that which it might have received had it been invested elsewhere. They will not, in consequence, abandon their property. So long as it yields them any net receipts over and above the expenses which must be incurred to keep the

place open, they will keep it open on the principle that some return is better than none. The reward received by the owners of the capital can only be determined as a surplus left over from total receipts after other factors have been rewarded at marginal productivity rates, and it is customary to call it a rent, although here it happens to be paid to the owners of capital.

On the other hand, if the promoters had under-estimated the demand, they might be able to charge higher prices than they had anticipated, and total receipts might be so large that when all other factors had been paid their market rates, the surplus was more than enough to pay a normal rate of return on the capital invested. The higher rate of reward would not immediately attract other promoters to build rival hotels and thus to increase the supply of accommodation. For some period of time it would be possible to regard the quantity of capital invested in hotel property in that place as fixed, and, so long as this condition held, the payments to the investors who had supplied the capital would be classified as a rent.

Units of any factor of production may find themselves in a similar position. Workers of a particular type may take several years to train. If the demand for their services increases, it will not immediately be possible to increase the supply. Then, if the supply of all other factors to this line of production is perfectly elastic, labour will appropriate the difference between the total payments to factors at marginal productivity rates and total receipts. From their nature, land and capital are more likely than labour to enjoy such casual surpluses and at the same time to suffer casual deficiencies of remuneration. In cases where the demand for a commodity, the supply of which is very inelastic, changes suddenly, the payments to all factors concerned will be in the nature of rents.

Quasi Rent

Marshall recognised the similarity between such cases and that of the rent of land, and termed the reward of a factor in these conditions a " quasi rent." He saw that when a producers' good is once in existence, it has to work for what it can get; its use cannot be withheld, as can free or floating capital. The return to an old investment of capital is therefore " more properly treated as a sort of rent—a quasi rent."

The reader may well be excused if by this time the various usages of the word rent have become somewhat confused, and a brief summary may be useful.

Summary

First of all there is in modern economics no special theory of the rent of land, for we no longer believe that land "has peculiarities of its own which are of vital importance from the point of view of theory as well as of practice." The rent of land is explained by means of marginal productivity analysis, in just the same way that the reward of any other factor is determined.

Economic rent, net rent, pure rent are names for a part of the actual payment made to land, viz. that portion of the actual rent which is a payment for the natural unimproved qualities of the soil. We deny the significance of this dichotomy, and reject the special theories which explain the phenomenon of net rent.

Finally, we have the various species of quasi rents, the hyphenated rents we may somewhat irreverently term them. We shall find scattered through economic writing such examples as scarcity-rent, entrepreneurs-rent, monopoly-rent and the like. They may be interpreted as signifying the earnings of units of some factor of production whose rate of remuneration cannot be determined on marginal productivity lines, because the quantity of factor employed there is fixed by the conditions of the problem.

Further Reading

BENHAM, F. C.: *Economics*, Chap. 18.
FETTER, F. A.: *Principles of Economics*, Part 2.
MARSHALL, A.: *Principles of Economics*, Book 6, Chap. 9.
TAUSSIG, F. W.: *Principles of Economics*, Vol. 2, Chaps. 42–44.

More Advanced Reading

ROBINSON, J.: *Imperfect Competition*, Chap. 8.

Chapter XIII

THE SUPPLY OF LABOUR

THE idea of discussing the market supply and price of human labour in the same manner as for commodities has appeared revolting to many writers on social problems, but if our approach to these problems is to be realistic, we must appreciate the necessity of some method of valuation for human services in any form of social organisation. Prices for labour of different kinds and qualities are essential to successful economic accounting, and it would be advisable to attempt to compute them even in a society which distributed income on a perfectly equalitarian basis.

If Jones's labour is worth only 25*s.* a week, it is important, both to a capitalist employer and to a communist planning commission, that the fact should be known, otherwise people who are capable of making a bigger contribution to the communal income may be retained in places where they are making a smaller one, and the income of the community is less than it might be. Also if Jones is not capable of producing a more valuable output, the fact should be known so that the reasons for it may be discovered. But the fact that Jones's labour is worth only 25*s.* a week does not involve the implication that it is socially desirable that he should have to keep his family on that income. That is a question for the social conscience of the community. It is true that interested parties have been in the habit of ascribing Jones's position to the working of an inexorable economic law, which makes it inevitable that somebody shall be in that case, and that misconceptions about economic theory have led many to believe such a palpable falsehood. Such a state of affairs makes a candid objective examination of the problem of wage determination the more essential.

Definitions

First it is desirable that terms which are used vaguely in common speech be more precisely defined. The term labour will be taken to include all forms of human service for which a payment is made, whether directly in the form of a wage, or in the form of an item in the computed cost of some article which is sold. So whether an individual buys leather and makes boots for sale, or whether he works for a wage on material belonging to another, there will be a payment to each in respect of the labour which has gone to make the pair of boots, and this we shall call a wage. The first individual may also draw wages in other respects; for instance, in his capacity as manager of the shop in which the boots are sold.

We shall not include as labour those services which each of us provides to himself as a consumer. No useful purpose is served in computing the wage I earn by shaving myself daily. Neither is the work done by women in their own homes included in the category of labour; not because it is not an important contribution to the real income of the community, but because the task of computation is a difficult one, and if the value were computed, it would not enter into our economic calculations. The result of our calculations is affected indirectly when a serious diminution of these services takes place, as when married women take up paid work in large numbers; but the effect on the real income of the nation cannot be computed in money terms.

All payments for services we shall call *wages*, even when the recipient prefers to call it a salary. As we have seen, a proportion of the income of an individual who conducts a business he owns will be included under the term wages. Wages are paid for services rendered to oneself if those services affect a business account, but not if they merely affect the private account of the performer.

At times it will be necessary to distinguish between *money wages* and *real wages*; the latter denotes the former turned into goods and services and serves as a measurement of wages, which is independent of changes in the purchasing power of money.

A *wage-rate* is the money payment per unit of time as laid

down in the wage contract, and the unit may be an hour, a day, a week, a month or a year; or payment may be per unit of output. In the first case it is called a *time-rate* and in the latter a *piece-rate*. Sometimes a wage contract may involve a mixture of the two. There may be a piece-rate with a provision for a minimum payment per shift or per week.

Finally, we have *wage earnings*, or the total payments in money or in kind in respect of labour services over a period of time, usually a year. It is necessary to point out that earnings may differ from the sum given by the wage-rate multiplied by the period of time, owing to the occurrence of unemployment, short time, sickness or other causes of stoppage. Also there may be deductions from wages, as so computed, for materials consumed, as when miners were charged for blasting powder.

The Supply of Labour

The problem of the supply of labour, like that of land, involves the same two questions of total supply and distribution between the various uses. One aspect of the supply of labour is peculiar to that factor alone: the supply is not merely a question of the number of workers, but also of the amount of work each of them is able and willing to do. We will consider each of these aspects in turn.

The Total Number of Workers

The problem of the size of the labour force of a community involves first of all the relation between the size of the population and the wage level. The relationship may be dual in character, that is, the size of the population may affect the average level of income, and the average level of income may affect the growth of the population. The discussion of these two problems constitutes what is generally known as the theory of population.

The Optimum Theory of Population

The optimum theory of population which examines the effects of size of population on income per head is a simple application of the Law of Diminishing Returns. It states that, given the natural and acquired resources of an environment, given the store of knowledge, skill and technique, there is one size of

population which will produce a larger output per head than any larger or smaller number. It is clearly a theory of the size of the working population, for it is the working population and not the total population which is relevant to the size of output. As the number of workers increases, all the advantages of division of labour are rendered possible by the expanding market and the expanding labour supply jointly. After a time these advantages are counteracted by the effect of the diminishing quantity of capital goods and other resources per worker, for it is a condition of the problem that all productive resources apart from the supply of labour remain constant.

The concept of an optimum population was considered by Cannan to offer a criterion for judging the state of a given population. If a country was over-populated, a fall in its population would lead to an increase in income per head. If it were under-populated, an increase in population would lead to a rise in income per head. In fact, the theory has proved useless for this purpose. Other things do not remain constant, and it is possible that population has to grow before a new technique can be adopted. Also the immediate result of a decrease in population would probably be a decrease in total output. This might come about even if the decrease were due to the emigration of unemployed. They were consumers, and provision for their consumption provided somebody with an occupation. Thus the departure of a large body of unemployed persons might well cause others to be unemployed. The reasons for this will be clearer after we have discussed the influences which determine the size of total income and employment.

There is indeed no way of determining statistically the existing state in relation to the optimum. The existence of unemployment is no proof of over-population. Surpluses of labour exist in certain trades at times when there are shortages in others. An apparent labour bottleneck may in fact be due to a shortage of a particular kind of machine, which, if released, would lead to a labour surplus in the trade where a shortage now exists.

In a very simple economy it may be possible to say whether over-population exists. In a subsistence agriculture economy, where the holdings of land have been subdivided until the

typical holding provides a bare subsistence for a household, the fact of over-population is clear enough. It may be possible to say that a given area is over-populated by a nomadic people, but that a much greater population could be supported if the land were irrigated and intensively cultivated. When it comes to a complicated economy like our own, a clear answer is probably impossible. As we have already seen, the short-run effect of a decline might be in the direction of a decrease in output per head, and yet the longer-run considerations might demand an encouragement of emigration. Thus, if Britain were to concentrate industrial effort on the mass production of common consumption goods and other countries were to follow the policy of making themselves self-sufficient in this respect, the emigration of a large part of the present population of these islands would be inevitable. But if a phase of capital construction in the tropical countries, such as occurred in the temperate countries last century, were to begin, the large quantity of skilled labour in Britain would enable a considerable advance in the standard of living to be made, possibly with an increase in population. As soon, then, as an attempt is made to apply the optimum theory to reality, it is necessary to step outside the conditions of the theory and to speculate on the probable factors of change during the next quarter-century or so.

Effect of Income on Population Growth

We have been unable to say anything very positive on the effect of the size of population on income per head, and we shall find ourselves in an even worse case over the inverse problem of the effect of the income level on population growth.

Through most of the nineteenth century this was a matter on which most writers held very positive opinions. The theory propounded by Malthus, that a rise in the general level of wages increases the rate of growth of population, was generally accepted. It was believed that human fertility rendered it inevitable that the production of an extra bite of subsistence would always lead to the emergence and survival of a mouth to eat it. The only checks on the growth of a population were misery and vice. If, for any reason, whether a technical innovation, increased skill or discovery of new natural re-

sources, the production of a community grew, the population would automatically increase until the increased production would once again provide a bare subsistence level for all. Improvement of social standards was therefore impossible, because the improvements would be swept away by an increase in births. If a Utopia should come into existence, it would not last long, as it would soon become overcrowded.

Malthus was generalising from experience. He pointed out that population had doubled in twenty years, and that if subsistence could increase by as much in the next twenty years as it had done in the past twenty, it would be a remarkable achievement. So, with population increasing in geometrical progression and subsistence in arithmetical progression, starvation and penury were the inevitable result. Where he went wrong was in assuming that the ratio in the geometrical progression was a constant one, and this was probably because he mistook the nature of the increase in population he was observing. It was due, not as he imagined to an increase in births, but to a decrease in deaths, particularly in the deaths of babies and small children.

A number of factors brought this about: the general improvements in sanitation and water-supply; the improvements in medical science, particularly the spread of the practice of vaccination; the production of cheaper clothing, which began to restrict the practice of passing on heavy clothing which was difficult to clean. All these reduced the incidence of typhoid, cholera and smallpox, which killed large numbers annually, and infant mortality was also reduced. This was a process which could not go on for ever, so that the growth of population would slow down even if birth-rate remained the same.

But in Western Europe and North America the birth-rate did not remain the same: it declined. The most important factor in this decline appears to have been the adoption of new and more efficient methods of birth-control in the last quarter of the nineteenth century. Many theories have been propounded about the effects of an increased standard of living on human fertility, but the statistical evidence is lacking. It is, for instance, a popular idea that the higher the income of a family the smaller the number of children per family. There is, so far as the writer is aware, only one investigation, done in

Scandinavia, which produces any facts relevant to this theory, and that enquiry showed no relationship to exist between the income of a family and the number of children.[1]

If by chance the proposition should be true, a variety of explanations is possible. Improved standards of living may cause a decline in biological fertility, or families which are relatively infertile become richer through the concentration of a large number of small fortunes in one generation in a smaller number of somewhat larger fortunes in the next generation, or a family may restrict the number of children in order to concentrate effort on the business of making money. Also it appears probable that in Britain the new birth-control practices originally affected births in the top half of the income scale much more than in the lower half.

Although the causes of the decline in birth-rates among people of Western European origin could be biological, the phenomenon could also be explained on the simpler hypothesis that it is a matter of deliberate choice. A variety of reasons may have accounted for that choice: the indifference which until recent years characterised the attitude of the medical profession to the painfulness of childbirth; the increasing cost of house accommodation and domestic labour relative to incomes; the expensive character of private provision, and the poor quality of public provision for care in maternity; the costliness in relation to average level of income of secondary and higher education, so that the number of children is limited to give the smaller number a better chance; the growing sense of the instability of the social structure in the twentieth-century world. In one of these factors probably lies the prime cause of the decline in birth-rate in this country in the past seventy years.

Malthus' Bogy not a Myth

So far as the white race is concerned, the devil raised by Malthus seems to have been in chains for a long time, but periodically those chains get a severe rattling. Despite the substantial increase in agricultural production, which has far exceeded the limits Malthus believed possible, it is nevertheless true that the world has not been producing sufficient food to

[1] Edin, quoted by Enid Charles, *Menace of Under-population*

provide the minimum standard necessary for physical efficiency for all its people. Those who have enjoyed the necessary minimum or something better have done so at the expense of others who were below it. Even in the 1930's, when food-stuffs were being destroyed because they could not be sold, a large proportion of the people of the world were undernourished, and no very large diminution was necessary to turn malnutrition into famine. According to Lord Boyd Orr, the world requires a minimum of twice the output of primary food-stuffs produced in the 1930's to provide a physiological minimum standard for all its inhabitants. If the raising of standards of living which this increase in food supplies would bring about should be accompanied by the same diminution in death-rates which occurred in Western Europe, a very substantial increase in world population would take place. It is not impossible that a further doubling of food supplies would be necessary before the decline in the birth-rate, resulting from the rise in standards of living, should ease the pressure on world agriculture. The Malthusian bogy is not a myth.

The Working Population

Not all the members of a population are available for work. Children below the legal or conventional minimum working age for the community, persons too aged or infirm to work, must be excluded; people who, because they have large incomes from property or who at the other end of the income scale are regarded as outcasts from society, refuse to work. We have already excluded women who work in their own homes from the labour supply.

The size of these categories of what we may term the voluntarily unemployed depends on the social structure and the custom and traditions of the community in question, and these must be known before any useful generalisations can be made. In Britain a rise in the level of real wages might be expected to decrease the number of married women who earn a money wage. In the lower income groups from which the mass of married female labour comes, a rise in the wage of the man is likely to be spent, partially at least, in reducing the amount of paid work done by the woman to enable the standard of comfort in the home to be raised. On the other hand, it is feasible

that, among recipients of so-called middle-class incomes, women would be more anxious to continue to work after marriage. The effect on the choice of unmarried women of working for a living or being supported by their fathers is much more debatable, but on balance there is little doubt that the numerical preponderance of women in the lower income groups giving up paid work to look after their homes means that the general effect of a rise in the level of wages is to decrease the amount of female labour. A rise in communal income increases the demand for cheap female labour and decreases the supply of it.

Similarly, it would be expected that, as family incomes rose, the period of education of children would be prolonged in order to fit them for higher-paid posts. It might also be expected to lead to children being apprenticed in trades where the eventual income was relatively higher, instead of being sent into occupations where they could make the maximum immediate contribution to family income irrespective of future prospects.

The age of retirement would probably be reduced by a rise in the level of real wages, which would make provision for old age possible. On the other hand, a decrease in invalidism would mean a reduction in the numbers whose working lives are cut short by ill-health, and also a reduction in the numbers of working days lost by all workers from that cause. The factors involved in the determination of the size of labour force a given population will provide are largely questions of custom and habit, and although custom and habit may have their origins in economic conditions, they are slow to react to changes in those conditions.

Taking the size of the working population as given, the amount of effort it will put forward at any given wage level is dependent on the ability and willingness of the population to work.

Effects of the Age Composition of the Population

Ability to work is not the same throughout the working life of an individual. There is an initial period of increasing efficiency when the worker is learning his trade and has not reached physical or mental maturity. Then, when maturity is reached, there follows a period during which the worker is at the height of his powers, and finally comes a period of decline.

The age of reaching maturity, the duration of the prime of life, and the seriousness of the eventual decline may vary from one race to another and from one occupation to another, and the conditions under which the population has to live and work will largely determine whether workers are worn out at forty or comparatively virile at sixty.

Consequently, if we have two populations with similar hereditary characteristics and living in similar environments but different in age composition, differences in output per head from this cause may be anticipated. Such considerations are most important when an attempt is being made to estimate the effects of a change in the size of a population. A population which is growing by natural increase, that is, by excess of births over deaths, will have a greater proportion of young children under working age, and consequently a smaller total income, than a population of the same size which is growing by immigration. Immigrants are usually young people approaching or in the prime of life, and a population which is receiving them will have an abnormally large proportion of them and the country losing them a relatively small proportion. Income per head will tend to be higher in the first case and lower in the second. A population which is declining through the failure of births to replace deaths will have a preponderating number of old people whose working life is over or who are in the phase of decreasing efficiency.

Ability to Work

Apart from differences in industrial efficiency arising from differences in the age composition of communities, there are still further divergencies to be investigated in the amount of work done per head. All observed differences in quantity or quality do not arise from differences in ability; willingness must also be taken into account, but it is also quite clear that ability is not everywhere the same. It is probable that questions of both inherited and acquired characteristics are involved, though it may often be difficult to decide to which type of cause a particular phenomenon must be attributed. For instance, there does not appear to be sufficient evidence to say whether the heavy labour turnover which has been a characteristic of Russian industry from the time of Peter the Great to the

present day is due to some personal characteristic inherited from nomadic ancestors, or to some feature of the Russian social or economic structure. However, our concern is rather to note differences than to indulge in, what must be for us, amateur speculations as to their precise causes.

Without giving any countenance to the gross absurdities which for political motives have been made to masquerade in certain European countries as scientific studies of racial problems, we may note that differences in mental and physical attributes exist between typical members of different nations, but that these differences may be due to differences in physical environment, in innate qualities or in training. The considerable differences which exist between members of the same community and which are directly attributable to differences in the adequacy of diet, clothing and shelter, must force us to give considerable weight to these environmental factors in accounting for differences between members of distinct communities.

A study of the workers of different races reveals differences in physical strength which may or may not be accompanied by similar differences in resistance to fatigue, manual dexterity, mental quality, and in powers of concentration, whether at the very low level of intellectual ability which may suffice for a machine-minder, or at higher levels. These factors will determine whether a given population can supply only manual labour of a type which requires nothing but physical strength; they decide the degree of complexity of the machinery labour may be given to use, and the quality of the work done, both from the point of view of delicacy of materials and design and also from that of freedom from faults.

The marginal productivities of equal quantities of labour from different races may therefore differ greatly, and in consequence differences in wages cannot be taken as reflecting differences in labour costs.

Inside a given community differences in industrial quality are attributable principally to the distribution of incomes in the preceding generation. If children have inadequate food and clothing, they can make but little use of even an elementary school education. A home in which there is space for privacy and quiet enables mental development to take place, and this

is impossible when a family is living in one room. The general efficiency of the population may also be raised by State expenditure on technical training, by means of which the marginal productivity of labour as a whole may be raised.

Demand for Income in Terms of Effort

At least as important as differences in ability to work are differences in willingness to work. From the point of view of the individual concerned, the quantity of labour he is willing to supply is the price he is willing to pay for income, and willingness to supply labour may be better regarded as demand for income in terms of effort. The individual may be considered as seeking an equilibrium between money income and the other uses of his time and efforts. In the case of primitive peoples, the other uses will include the provision of food and perhaps clothing. The quantity of labour supplied will be related to the demand for money to pay taxes and to purchase a few European goods. With industrial workers who are dependent on their money earnings to purchase everything they consume, the choice is between the goods and services money income will buy and the leisure which must be sacrificed to earn it. It is a question of the alternative uses for time and effort, and an equilibrium is reached when the increment of goods and services an extra hour's work will secure is valued equally with the leisure which has to be sacrificed to obtain it.

At first sight it might appear that the demand for income is unlimited, and so it is in the sense that the demand for anything may be unlimited if it can be had for nothing, but income must be paid for in time and effort, which have alternative uses. The overtired worker may be incapable of using his non-working hours for anything but eating and sleeping, and a reduction in the total amount of effort required for work, which would enable him to develop some hobby, may be more important than an increment to even a very small income.

It is convenient to consider the problem in terms of hours of work and hours of leisure, assuming that the expenditure of effort per hour remains the same. It is merely a matter of convenient symbolism, for the analysis is equally true if worked out in terms of effort measured in foot-pounds.

It is a common practice to assume that the response in the amount of work done to a change in wages is invariably in one direction. People who invent complicated formulæ which they allege will calculate the wage-rate any particular type of labour should be paid, appear to assume that a higher wage-rate invariably calls forth a greater supply of work per employee. On the other hand, many writers on public finance make the contrary assumption. If this were indeed the case, the task of the Chancellor of the Exchequer would be an easier one. The higher he raised the income-tax, thus reducing the real wages of everybody, the harder everybody would work and the greater the output of the community would be. Both generalisations are unsafe, as is readily seen when the problem is restated in the form of the demand for income in terms of effort.

The supply of labour is the payment for income, and the effort price of income is the total amount of work performed divided by the amount of income received. A high effort price, therefore, corresponds to a low wage level and vice versa. In Fig. 36 effort prices are measured on the vertical axis and quantities of income on the horizontal axis. The demand for income in terms of effort may therefore be represented by means of a normal demand curve. The total output of effort put forward to earn a given income will be represented by the area enclosed by the co-ordinates. The assumption that the amount of effort put forward increases as the wage rises or the effort price of income falls, means that the demand for income in terms of effort is invariably elastic, and the contrary assumption requires elasticity to be invariably less than unity. It is highly improbable that either of these propositions holds true of everybody, though the assumption of inelastic demand for income in terms of effort may well be true of a larger proportion

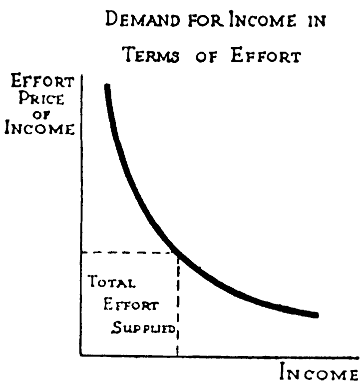


FIG. 36.

of the population in this country than the contrary proposition. The recipients of low incomes are perhaps likely on balance to put forward more effort in an attempt to prevent a reduction in income, for their consumption is already reduced to the barest necessities. On the other hand, a rise in wage-rates will cause them to take out some of the possible increase in income by a relaxation of effort. The low level of income to which they have been accustomed will have given them little experience of consumption outside the bare necessities of life, and once these are covered, the marginal significance of further increments of money income declines rapidly. This is another aspect of what has been termed "the terrible wantlessness of the very poor."

A more extreme case is that of the primitive who, when he has purchased what the white traders have made him believe to be conventional necessities and has saved enough to pay his taxes, refuses, with complete rationality, to engage in further paid labour.

Generalisations as to the elasticity of demand for income must be based on observation of the labour groups to which they are to be applied, and on this point, as elsewhere in questions of the supply of labour, there is much the social anthropologist might do to make data available for use in economic analysis.

The Scope for Individual Choice

It may be objected that the idea of finding an equilibrium between income and leisure may have some relevance to the case of piece-rate workers, although they cannot vary the number of hours they shall work, but that it cannot apply to time-rate workers, whose working day is fixed and who earn no more however hard they work. The individual worker has presumably used his vote in determining the contract arranged between the union and his employers. The effect of the intervention of the union is illustrated by the problem of overtime, which most unions try to reduce to a minimum. Without a union, the worker will work overtime, even though he realises that to do so is injurious to his health, and by reducing the number of jobs impairs his chance of getting another if he loses the one he has. Refusal to work overtime is indeed likely

to mean dismissal. The existence of a union enables him to take a long-run view of his own interest, and it is only from the very short-run point of view that income is lower than it otherwise would be.

In the case of much black-coated labour, it is possible to take work home out of office hours. No extra pay is earned for this, but it is advisable to do it, if promotion and the higher income it brings are desired. Almost any large office will contain at least one individual whose work is always efficiently performed, but who will never receive promotion and who indeed will never seek it. He regards his occupation as providing him with the means to follow some pursuit in which he is really interested but which cannot procure him a living. The scope for choice is indeed wider than may at first appear.

The Supply of Labour to Particular Occupations

The general principle governing the distribution of the labour supply between its alternative uses we have already discovered in the tendency of units of a factor to move so as to equalise remuneration in all uses. But having stated the principle, it is at once necessary to point out that it is a statement of tendency only, and that in many cases the barriers are such as to prevent equality ever resulting. The barriers to migration across national frontiers are such that important wage differences can exist permanently. Even in the absence of a law restricting immigration of labour, language and cultural difference between the immigrant and the people among whom he settles are sufficient to deter movement. Nevertheless, there is the case of the rapid growth by immigration of the population of the United States, where eventually immigration restriction was introduced to prevent depression of the wage level. Despite the popular view of the United States as the country with the world's highest wage level, immigration of low-grade Eastern European labour had lowered the wage available to this type of labour to levels which were certainly not generous on Western European standards.

Similar differences exist between different districts of the same country, and, even in a country as small as Britain, differences in speech, diet, housing and habits of living generally are large enough seriously to reduce mobility between districts.

Mobility of Labour

Experience between the wars showed that considerable unemployment may exist even when most industries are prosperous and when some are suffering from a shortage of labour. This has been on an entirely different scale from the "reserve of labour" amounting to 1 or 2 per cent. of trade-union membership in the nineteenth century, and which, existing in times of boom, consisted of workers temporarily unemployed between jobs. It appears now that large-scale unemployment existed in the United States in the boom years of 1928–29, and in the 1930's one spoke of an irreducible minimum in Britain of 10 per cent. of the workers included under the unemployment insurance scheme. Such bodies of unemployed largely consist of people who are displaced owing to a shrinkage in demand for their products, and in Britain were drawn largely from the cotton and coal-mining industries, where such decreases in demand had taken place. At times when unemployment among these groups was very high, there was a shortage of bricklayers. Why, then, could the unemployed not fill the vacancies?

One of the reasons was that trade-union regulations frequently prevent adult workers who have once been attached to one trade from learning another. The attitude may appear heartless, but the unions have had so much to fear in the past from employers who diluted skilled labour with low-paid unskilled workers that their attitude was understandable. When heavy unemployment is a constant feature of industrial life, each group of workers will endeavour to ensure that it gets the largest possible share of the employment which is available.

The retraining of middle-age men for new occupations is not one which employers will undertake when they can get young labour which is more easily trainable, the older men having lost the elasticity of youth. At the same time it is cheaper to train young people, because grown men cannot be employed at boys' wages. Further, it requires a very long period of unemployment to convince a man that the demand for the services he has performed in the past has disappeared, and that if he wishes to be employed it must be in another occupation, probably one which in his eyes is lower down in the social scale. The period

required may be so long that the worker has become unemployable in the process. This kind of consideration applies chiefly to occupations where a very specialised skill is important. A great deal of industrial labour is skilled at minding machines and can fairly quickly master the peculiarities of another type of machine.

The cause of much of this chronic unemployment is to be found in the concentration of industries in certain localities. If there is but one industry in a district and that industry meets a fall in demand for its products, then a change of job means a move to another district, and there are considerable resistances to such movement. Attachment to local habits of life, differences in speech, fears of being stranded among strangers and bereft of the social contacts through which help in getting another job might be obtained, all tend to tie people down to the familiar locality.

If such movements are to be organised, it is necessary to move family groups, and greater success has been achieved when the units have been still larger. As a national policy, transference of labour on a large scale has the disadvantage that housing and ancillary services cannot be moved with the people: amenities are thrown idle in the places from which labour moves and have to be provided in the places to which they move. Big movements of population have of course taken place inside this country, but these have taken place as the result of a series of short-distance movements, so that few individuals move very far. When the movements which fed the concentrations of population in the Midlands of England took place, people moved in from surrounding counties and others moved to take their places.

Allocation of New Entrants to Industry

Just as changes in densities of population can come about as the net effect of many short-distance movements, so industries can wax and wane without individuals changing their jobs. If the supply of young labour can be cut off from an industry, it will shrink through failure to replace deaths and retirements. A growing industry, like a growing population, has an abnormally large proportion of young people and a declining industry an abnormal proportion of elderly people. Unfortunately,

young labour too often continues to pour into declining industries.

Important causes are ignorance of opportunities elsewhere and the short-sightedness of parents in sending young people to work in industries where there is a probability they will become unemployed as soon as they reach an age which the trade board lays down as qualifying for adult wages. Again, the worst effects have been seen in districts with but one industry, where a young person wishing to enter another type of industry could only do so by leaving home, and earnings were not sufficient for support outside the family circle.

Social Stratification

Even if it is not true that certain occupations are reserved for those who have been educated at certain schools and certain universities, yet it cannot be denied that the possession of the correct educational label both smooths and shortens the path. In Britain in the past the social class in which a child was born has been more important from the point of view of access to higher-paid posts than either education or ability.

Such obstacles exist at every grade; they make it difficult for the son of a bank clerk to become an ambassador and for the son of a manual labourer to become a doctor.

Training Costs

Apart from the difficulties arising from social snobbery, there are also the formidable difficulties of costs of training. The existing distribution of incomes has considerable influence in deciding the number of entrants to different occupations, although the provision of free and assisted education has done something to remedy the effects of unequal distribution of incomes. One of its principal achievements has been to break the monopoly of "black-coated" labour enjoyed by one social group in the nineteenth century, but much more remains to be done before anything like "la carrière ouverte aux talents" exists.

The cost of training limits supply, and so, if the conditions of demand which made it appear worth-while to incur the cost of training persist, the income achieved will be high. The level of income ultimately attainable is not, however, the sole criterion; the non-monetary attractions of an occupation are

also considered. The experiences offered by an occupation, the social prestige attaching to it, the security of tenure it offers, the leisure to pursue other interests which will not afford a means of livelihood—all may influence choice. In this connection it must be remembered that it is not the individual who receives the training who usually pays for it. Then again it is not the general level of income which is always taken into account; the existence of a few particularly rich prizes may attract, although the general level of remuneration is low. Thus, the Bar probably offers the highest earned incomes of any occupation in this country, but the number of fashionable counsel earning high fees is small in relation to the large number of barristers who have to find other ways of supplementing income.

Non-competing Labour Groups

The picture of the labour supply which emerges from this discussion of some of its characteristics is that of a highly stratified supply. Individuals tend to find themselves forced by circumstances into a particular stratum. Movement from one stratum to a more highly paid one is difficult once an individual has been allocated to a particular occupation. It is far from easy for a child of parents in one stratum to climb to a higher-paid one, so that there is little competition upwards through the strata, despite the fact that many individuals do succeed in climbing.

Further Reading

Cannan, E.: *Wealth*, Ch. 4.
Hicks, J. R.: *Social Framework*, Part 2.

More Advanced Reading

Carr-Saunders, A. M.: *World Population.*
Carr-Saunders and Caradog Jones: *Social Structure of England and Wales.*
Charles, Enid: *Menace of Under-population.*
Florence Sargant, P.: *Labour.*
Glass, D.: *Population, Policies and Movements.*
Hilton, J.: *Are Trade Unions Obstructive?*
Marshall, T. H., and others: *The Population Problem.*
Reddaway, W.: *Economics of a Declining Population.*
Walshaw: *Migration to and from the British Isles.*

Chapter XIV

THE THEORY OF WAGES

HAVING discussed the causes which determine the total supply of labour and those which stratify the total supply into sectional supplies between which mobility is imperfect, it is now appropriate to examine the conditions of demand for labour, and proceed to the determination of wages as the price which equates supply and demand.

We are not, however, ready to discuss the problem of the demand for labour as a whole, that is, to discuss the theory of employment. The total number employed depends on the demand for commodities which is in turn related to income, and we must discuss the rôle of capital in the productive process before we can turn to income.

If we assume that the income of the community is fixed, we can usefully discuss the demand for labour of particular kinds and the determination of its wages, and so fill in one of the gaps in our knowledge of the costs which figured in our study of the pricing of commodities.

The Marginal Productivity Theory of Wages

In Chapter VII we discovered that the price of a factor in production is equal to the value of the marginal physical product, which, under competitive conditions, is equal to the marginal physical product multiplied by the price of the product. This comes about because the buyer of a factor is willing to increase his use of that factor so long as the increment in product obtained by using one more unit is worth more than the sum which has to be paid for the use of that unit. When this condition ceases to hold, the use of that factor is no longer extended.

For labour, then, this is to say that a given employer will increase his employment of labour of a given kind so long as the

extra output produced by an additional worker is worth more than the worker's wage. The demand schedule of a particular firm for a given kind of labour will therefore consist of a series of pairs of quantities, one of which is a money wage and the other that quantity of labour whose marginal physical product when employed by that firm, multiplied by the price of the product being made, is equal to the wage. The market schedule for this kind of labour will be found by adding the quantities of labour each employer will demand at the given wage level. The demand curve will be part of the downward sloping portion of the curve of values of marginal products for all users combined.

If we are to be able to use such an analysis of the wage-fixing process, it is necessary that we shall have groups of units of labour which are reasonably homogeneous, and we have already seen that the supply of labour is divided into a large number of sectional supplies or pools of labour. There are commonly differences of kind and quality between different pools of labour, but the workers who make up a given pool are providing the same type of labour, even though some differences of quality may exist.

Furthermore, these pools of labour have come into existence because of the sectionalisation of demand. People have taken the necessary steps to become navvies or physicians, cotton spinners or bricklayers, in a particular locality because *at the time they decided to follow that occupation there was a demand* for that type of service in that locality. It is therefore not possible to talk of either demand or supply of labour as a whole, but of demands and supplies of particular kinds of labour. The demand schedule of an individual employer for labour of a particular type is the descending portion of the scale of values of marginal products, so that the downward sloping portion of the curve of marginal value products is the demand curve for labour for an individual employer. By adding together individual demands, we can obtain the total demand schedule for labour of a particular type.

The considerations of the supply of labour we have discussed in Chapter XIII suggest that the supply curve for labour will be of the normal ascending type. The number of people in a given locality who call themselves, say, bricklayers is ascertain-

able at a given time, but that does not mean that the supply of bricklayers is fixed irrespective of the wage paid. At certain levels of wages some members of the pool may prefer to work at some alternative occupation, to go to another district, or even to be unemployed, while at higher levels of wages persons who were once bricklayers may see fit to return to their former occupation and bricklayers from other districts may consider it worth-while to move in. There is also the possibility of variation in the amount of work done per man as the wage level changes, but, as we have seen, the effect may be for a rise in wages to either increase or decrease the amount of work done.

In a competitive market for labour, the wage will tend to settle at the common point of the two schedules. If the demand and supply schedules overlap at all, there must be one wage level at which all who are willing to work for that wage are employed and all who are willing to buy labour at that price are satisfied. This is the equilibrium wage in the conditions of the market, and it will depend on the particular conditions of demand and supply in the market for the type of labour in question.

Problem of Women's Wages

The problem of the differences in the wages paid to male and female labour is one which serves admirably to illustrate the principles we have been investigating. Men and women can be regarded as two groups of labour, between which mobility is zero. For some purposes male labour is undoubtedly superior to female, and for others the reverse is the case, but in many occupations there is no obvious advantage either way. It is to be noted, however, that women's wages, in occupations which are exclusively female, are lower than those in male occupations where a comparable degree of skill, training or general education is demanded. Custom and prejudice play a considerable part in determining how far occupations shall be open to women, and these are still strong enough seriously to restrict the field of female employment. The most intense prejudice attaches to the employment of women in the higher-paid technical or administrative positions, while no compunction is felt about employing them exclusively in semi-skilled or unskilled occupations for a wage no male worker would accept.

Since female labour is allowed to compete seriously with male only at the lower end of the male wage scale, the field of employment for women is small relative to the number of women seeking employment, in spite of the fact that the number of women who wish to be employed is less than the number of men. In consequence, the marginal productivity of female labour is lower than that of male labour, a result which is independent of the question whether the efficiency of a woman is, or is not, equal to that of a man. The lower marginal productivity must follow from the fact that the field of employment is more restricted relative to numbers, so in order that all the women who wish to be employed shall find jobs, it is necessary to move farther down the scales of diminishing marginal products in those occupations which are open to women.

Equality of remuneration as between the sexes can only be achieved by the removal of barriers, whether of law, custom or prejudice. The result of throwing open new occupations to women after the first world war was seen in the considerable rise in the wages of the lowest-paid grade of female labour, domestic servants, and the further extension of the field of female employment after the second world war led to a still more drastic diminution of the supply of paid domestic labour.

The organisations which are striving to secure the removal of barriers to the employment of women are thus working for the improvement of women's wages as a whole, but their campaign for equality of pay in those occupations which are open to both sexes will not invariably have this effect. In those occupations where freedom of entry is little more than nominal and where labour is wholly or mainly male in spite of the alleged removal of barriers, the use of female labour is an experiment from the employer's point of view, and there will be little inducement for him to experiment if there is no advantage to be derived if the experiment is successful. Equality can therefore mean that women are not employed.

The attitude of the male trade unions, which insist on women being paid at the same rates as men, is therefore quite logical from the point of view of their own interests. Women are excluded, and an increased supply of labour, reducing marginal productivity and lowering wages in that occupation, is avoided. The white skilled-workers' trade unions in South Africa found

the same device useful for excluding coloured workers whose competition might have lowered wage-rates.

In other occupations, like that of school-teaching, where women are in the majority, the enforcement of equality would be likely to exclude the men. A higher wage is paid to men because in certain circumstances there is a preference for them, and it is wished to maintain a supply for that purpose. Equality is likely to be achieved only by levelling down men's wages to women's, and this, presumably, means the exclusion of men.

These conclusions must not be taken to imply that there is no case for equality on the grounds of social justice, or that the present division of occupations between the sexes is likely to be permanent. It does appear that the problem is rather more complicated than is frequently assumed.

Wages and Numbers Employed

The assumptions under which we are working lead to the idea of a rigid demand schedule for labour. At any wage W a given employer will employ x workers of the type in question, because when he does so the marginal physical product of x workers valued at the market price p for the product is worth W. If we consider the employer to be selling the product in a competitive market, he can sell any quantity he chooses to make at the market price p, so that if the market price should change we shall require a new schedule.

But what if the price of labour should change, rising say from W to W_1? The employer will be forced to restrict his use of labour until the marginal physical product of the smaller number x_1 employed, valued at the same market price p, is equal to the new wage W_1. The change in the price of labour will, however, have upset the previously existing least-cost combination of factors. The rise in the cost of labour will have rendered it cheaper to produce an extra unit of product by using more of other factors and less labour. The proportion of labour in the combination will therefore be reduced, and the proportion of other factors increased.

If all other makers of the same commodity are affected in the same way, the demand for labour to make this commodity will be reduced, and labour will be thrown out of employment so far as the manufacture of this commodity is concerned. The

final effect will depend on the causes which have led to the rise in wages. If there is an increased demand for this kind of labour to make some other commodity, then what has happened is part of the process of redistributing labour in accordance with the change in demand for commodities. It will cost more to make an increment of the first commodity, since the marginal physical product of the factors other than labour will decrease as the proportion of them used is increased, so we may expect the price of the first commodity to rise also. This will bring the process of transfer of labour to a halt more quickly than would otherwise be the case.

On the other hand, if the wages of labour are arbitrarily raised, there is no compensating demand for labour to absorb those thrown out of employment. There are now several alternatives. Those who lose their jobs may prefer to remain unemployed rather than accept lower wages for doing the same job, or rather than detach themselves from the pool of labour to which they have previously belonged and attach themselves to another. Alternatively, they may decide to change their occupation, or to move elsewhere, or both. Finally, it is possible they may get back into their old occupation by accepting a lower wage. The same conditions would emerge if the demands for this kind of labour decreased for any reason whatever.

Under the assumptions we have made, there is thus no room for involuntary unemployment, except through frictions which prevent the second and third alternatives from operating. Consequently, studies of the causes of unemployment have paid a great deal of attention to the causes of immobility of labour, both on an occupational and a regional basis. Much unemployment can be accounted for by imperfections in the labour market: it was for this reason that employment exchanges and schemes for decasualisation of dock labour have been introduced. Important as these market imperfections are, they are not the sole nor even the major causes of unemployment, and the conclusion that was once based on them, i.e. that if all frictions immobilising labour and preventing wage-rates from being flexible were removed unemployment would fall to negligible proportions, is now shown to be erroneous. The little we can say here on this problem must await our discussion of the causes affecting the volume of employment as a whole.

Tendency to Uniformity of Wages

We have seen that the wage level inside a particular sectional labour market depends on conditions of demand and supply there. The relative levels of wages ruling in two different markets will therefore depend on differences in conditions of supply and demand. The wages of woollen weavers in the West Riding and in Gloucestershire will be the same only if conditions of supply and demand are the same. The demand for weavers in each area will depend on the demand of consumers for their products, and as the products are not the same, the conditions of demand may differ. The supply of weavers in each place will depend on demands for labour for other purposes, as these will determine whether weaving is a popular or unpopular choice for young people in the district. If there is little movement between the two districts, conditions of supply may be independent, but if movement is customary, conditions in the two areas will approximate and wage levels cannot differ greatly.

Again, within the same district there cannot be great differences in the wage levels of two similar occupational groups, as it is easy for an individual to move between them. If the type of ability and training required by two occupations are very similar, wage differences may nevertheless exist if they are sufficiently specialised to make movement between the two difficult, for supply of labour depends largely on what demand was at some time in the past and wages depend on what demand is now. The tendency to equality of remuneration for units of labour of similar kind and quality is dependent on movement between the occupational groups being free enough to unify the supply of labour.

Equality of Net Advantages

Even so, the tendency to equality is not to equality of money wages, but to equality of net advantages. Different occupations offer different conditions of non-monetary character. Some jobs offer work in pleasant surroundings, interesting experiences, security of tenure, ample leisure, freedom from detailed supervision; while others offer an unpleasant environment, dull routine, risks of accident and disease, insecurity and

small leisure. The individual makes his choice from the opportunities which are open to him, and takes into account the whole circumstances of the occupation. It is true that the most unpleasant jobs are often the worst paid, but jobs in which amenities are particularly good normally carry salaries which are somewhat less than others demanding similar training and ability, but offering less favourable conditions.

Nor is it always the average level of income in an occupation which decides choice. An occupation like literature or the Bar, which offers a few very rich prizes, may attract a supply of entrants which is quite unjustified by the expected level of earnings the average entrant can hope to achieve.

Wages with Employers' Monopoly

Let us now remove the assumption that conditions in the market for the commodity and the market for labour are approximately competitive. First, let us suppose some element of monopoly exists in the market for the product. Now the value of the marginal product of labour will no longer be the marginal physical product multiplied by the price of the product. Instead, it will be the marginal physical product multiplied by the marginal revenue from the sale of the product. As we have seen, marginal revenue is less than price, so the monopolist's demand price for any quantity of labour is less than if he were selling in a competitive market. If he is buying labour in a competitive market, he will therefore employ less of it than one of a number of competitive sellers would do.

It is also possible that a firm may enjoy some element of monopoly as a buyer of labour. Suppose, for instance, that a firm sites itself in a small country town where it is the only important employer of industrial labour. It will, like any other firm, equate the marginal cost of labour to it with the value of the marginal product, but it is not necessary that the marginal cost of labour will be equal to the wage-rate, as it is in a competitive labour market. It is indeed probable that the firm will be faced with a rising supply curve for labour. When it is employing all available industrial labour within a certain radius, it may have to pay higher wages or to provide free transport services to attract labour from a wider area. Marginal cost of labour would then be rising, and would be in

excess of average cost or of the wage-rate. This again will mean that the firm will employ less labour, produce less product, and sell at a higher price than would a similar-sized unit selling its product and buying its labour in competitive markets. Labour in these circumstances is commonly said to be exploited.

Trade Unions and Wage-rates

There may, of course, be monopolistic organisation on the side of labour, and it is possible, though not necessarily the case, that organised labour can exploit employers or more probably the consumer of its products.

If the employing firms are selling their products in a competitive market and are competing freely with each other for labour, a trade union may succeed in forcing them to pay higher wages, but it will not in general or for long be able to demand also that all its members shall be employed. If the employers' demand is sufficiently inelastic, it may be profitable for the members of the union to keep wages high and support a number of unemployed members, but the policy would soon cease to be profitable unless entry to the occupation was very strictly controlled.

Restriction of entry has been successfully practised, the medical profession being an outstanding example. There are, however, limits to the possibilities of restriction of entry. There must be complete control of the whole field of employment, and even then, if restriction is carried too far, pressure from members to get entry for friends and relatives may become irresistible.

The trade union has more scope where there is some element of monopoly on the employers' side. It may be a matter of forcing the employer to pass on some of the monopoly profit extracted from the consumer, or reducing the benefit the employer derives from his monopoly of employment and raising wages to the competitive level.

It is often claimed that a trade union forces employers to keep abreast of technical developments, and cases where this result might follow are undoubtedly possible. If a group of firms has a monopoly of a localised labour supply, it may continue to pay very low wages, which are nevertheless equal

to the value of the marginal product of labour in the technical process which is being used. If there is tacit agreement among the firms not to introduce new technical methods, which will probably require considerable new investment, the low wage level may persist during an upward trend of wages, until the occupation is regarded as "sweated labour." If a trade union is formed and is successful in raising wages, it makes the old method unprofitable and obliges employers to modernise their production methods. In general, it is claimed that the goad of trade-union activity forces employers to keep up-to-date in an endeavour to reduce labour costs.

Nevertheless, trade-union action may fail in its object, and the 1926 coal strike was an example of a very strong union breaking itself in an attempt to maintain a wage level that no exploitation of consumers, no confiscation of employers' profits and no technical changes which could be introduced in the time available, could succeed in paying. If trade-union leaders exercise foresight, they will keep wage-rates in line with the trend of value of marginal product, when all but a few per cent. of their members are employed, and pressing closely enough on rises in that quantity to ensure that employers endeavour to keep it rising. Their influence will then be in the direction of maximising the communal income. Policies which result in substantial unemployment and involve monopolistic exploitation of the consumer have the contrary effect.

Another set of trade-union practices becomes important when a country is deliberately following a policy of providing work for all who seek it. In the interests of securing their members against unemployment, unions have in the past secured agreements with employers reserving a certain quantity of employment for their members. These provisions include rules limiting overtime and setting out the kinds of work members of certain unions must and must not be asked to do. Technical changes may render these rules obsolete, and they hamper the mobility of labour, without which a policy of full employment may be very expensive.

Trade-union Activity and Equality of Payment

The market for labour is so sectionalised and so full of elements of monopoly that the tendency to equality of payment

to workers of the same type would hardly be very strong in the absence of trade unions. Indeed, trade unions probably have the effect of introducing equality of wages where differences in marginal productivities would otherwise exist. If markets for products and supplies of labour were highly localised, values of marginal products would differ according to local conditions. A trade union may enforce the same wage rates for specific jobs throughout the country.

Thus, if plumbers are more plentiful in relation to the demand for their services in one district than in another, there would be a tendency to lower wage-rates there, so that all should be employed at a wage equal to the value of the marginal product of the local supply. Then, if a trade union enforces equality of wage-rates everywhere, wages in such a district will be raised and the numbers employed will fall. In the short run there will be unemployment, but if the wage-rate does not exceed the average marginal product over the whole labour group, there will be opportunities for the displaced labour to be reabsorbed elsewhere. If, on the other hand, the rate is fixed at the highest paid anywhere, there will be unemployment among plumbers in all districts except those which formerly paid the top rate.

Again, trade-union activity tends to equalise wage-rates for workers of the same type but of different quality. Inside a particular occupation this may mean that some workers are unemployed because they are not worth employing at the full union rate, although they would be worth employing at a lower wage. Partially disabled persons who are capable of " light work " may be unemployed because uniformity rules prevent special arrangements. Again, maldistribution may occur when equality of rates is enforced for similar occupations which employ labour of the same kind but of different quality. It may be that in one occupation labour of superior skill and physique produces a greater product than lower-grade labour, but in another occupation there is little or no advantage. Total output will therefore be greater if there is a wage difference between the occupations, so that the higher-grade workers are attracted into the occupation where their marginal productivity is greater, and the lower-grade workers are employed in the other occupation at a wage equal to their lower marginal

STAFF C

productivity. Equality in these circumstances brings about a misdirection of labour resources and a reduction in national output.

Full Employment and Equality of Payment

The primary reason why a government adopts a policy of full employment is found in the belief that by central planning it is possible to bring about adjustment of demands and supplies of labour of particular kinds more rapidly than if the process is left to the pricing system. The essence of such a policy is that demands for labour shall be created to replace decreases in demands, however occurring. Supplies of labour of different kinds must be adjusted to these changing demands. From the very nature of a full employment policy, the old-fashioned method of forcing labour, through the privations of unemployment, to find itself new jobs is ruled out. This leaves two possible methods—direction of labour or manipulation of wage scales. Direction of labour is unlikely to prove acceptable, except in grave emergency, in the kind of society in which we live, so that adjustment of wage-rates appears to be the only method available.

It would seem, then, that a full employment policy might involve raising wage-rates in certain occupations in certain places without similar increases in similar occupations or the same occupations in other places. Similarly, if it became necessary to decrease supplies of labour in certain occupations and/or places, deliberate reductions in wage-rates, designed to make people change occupation or district, might be necessary. Trade-union uniformity rules would be obsolete in such a situation.

Differential Wages as Incentives to Effort

We have already discussed the effects of changes in wage-rates on the output of effort from a given group of people in terms of elasticity of demand for income in terms of effort.[1] There we considered the effects of a change in the wage-rate paid for each unit of effort. But just as in our study of discriminating monopoly we saw that it may be possible to induce people to make a bigger expenditure by charging different

[1] See above, p. 191.

prices for different units, so it may be possible to induce them to put forward a greater effort by allowing them to earn different units of income on different terms.

This is the idea which is behind one type of wage-bonus scheme. The general form of such a scheme is that a flat-rate basic wage is paid, which is the remuneration for the performance of a certain maximum task, and for output above this level a higher rate per unit is paid. The scheme may be based on collective or individual output. Thus, all the workers in a factory, a department or in a single gang may be paid a flat rate per week, so long as output does not exceed a determined figure. When output exceeds that figure, a bonus related to the excess is paid.

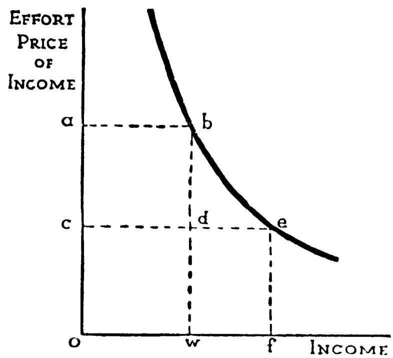


FIG. 37.

Let us examine the problem in detail for an individual who is paid a flat weekly rate w for an output of 100 units per week. When he is producing 100 units a week this is equivalent to a rate per unit of $\frac{w}{100}$. The effort price of a unit of income is then $\frac{100}{w}$. The curve in Fig. 37 represents the demand for income in terms of effort of this individual, *oa* represents an effort price $\frac{100}{w}$. At this price an income w is earned and an effort of 100 units put forth. We will assume that the demand for income is inelastic, so that at an effort price *oc*, which is half *oa* or $\frac{50}{w}$, an income *of* is earned for an output *ocef*, which is less than 100 units. Consequently, a straight raising of wages from $\frac{w}{100}$ to $\frac{w}{50}$ will result in a diminution of output.

But suppose that the individual is given the opportunity to earn more income at an effort price of $\frac{50}{w}$ provided he has already produced 100 units for a total wage of *ow*. Now we cannot say that our individual will be willing to earn an extra income *wf* for an extra output of effort *wdef*. He would be willing to do this if he happened to be earning *ow* for an effort of *ocdw*, but he is not. He is earning *ow* for twice that amount of effort. He would, nevertheless, be willing to take some increase in income at the effort price of $\frac{50}{w}$ per unit, and this output would be in addition to the output of 100 units.

The question of whether it would be profitable to the firm to pay $\frac{w}{50}$ for all units above 100 depends on the behaviour of marginal cost, assuming the product is being sold in a competitive market, so we can ignore marginal revenue. The extra bonus rate will mean that marginal labour cost is doubled, but this may be compensated by decreases in other of the component items of marginal cost. If the increase in output is made by working at greater speed, so that machines are more fully occupied, there may be no increase in cost from this source. Power, heating, lighting and supervision charges will not be increased. Cost of materials per unit may remain constant. It is possible, therefore, that marginal cost may have been reduced for these extra units, and output can profitably be increased until it is again equal to market price.

A bonus system may, therefore, induce workers to increase output, even when demand for income is inelastic. In the alternative form of bonus system which is perhaps more frequently encountered in Britain, the system gives a larger total wage but a lower unit piece-rate as output increases.

Wages and the Marginal Productivity Rule

We began this discussion of wages with the familiar rule that the payment to a factor of production is equal to the marginal physical product of the quantity employed at that price, multiplied by the price per unit of the product. We have seen that this rule is not of universal application, but that it is indeed

one instance of a more general rule, that a firm will increase its use of a factor until the marginal cost of that factor is equal to the increment in receipts due to the application of a further unit. The latter statement reduces to the first, when the firm is buying the factor and selling the product in a perfectly competitive market. The theory of wages consists of the application of this universal rule in any circumstances of supply and demand which may be encountered.

Further Reading

Bowley, A. L.: *Wages and Income since 1860.*
Dobb, M. H.: *Wages.*

More Advanced Reading

Douglas, P. H.: *Theory of Wages.*
Hicks, J. R.: *Theory of Wages.*
Isles, K. S.: *Wages Policy and the Price Level.*

Chapter XV

CAPITAL AND INCOME

In our discussion of the two groups of factors of production, Labour and Land, we were not aware of any difficulty in defining the factor. Labour is readily defined as consisting of all human services, but there was a certain vagueness over land, which we defined as the natural and acquired characteristics of the earth's surface. The vagueness arose from the difficulty of setting any limit on the acquired characteristics, short of all the constructional work men have accomplished. The older practice, of defining land as the original unimpaired properties of the soil, was an attempt to deal with this difficulty, but it failed to draw any useful line of distinction, because it is quite immaterial, from the point of view of any economic problem, whether some permanent characteristic of our environment is a natural phenomenon or whether it was created by human agency at some time in the past.

Definition of Capital

In defining capital, this problem of adapting the definition to the purpose for which it is intended to be used has proved even more difficult. If we wish to define capital in such terms that, together with land and labour, it completes the total of resources available to a society, then we should define it so that with land it covers all the characteristics of the physical environment. Some arbitrary division between those acquired characteristics which are land and those which are capital would have to be made, the line being drawn solely from the point of view of the usefulness of the definition in solving particular problems.

This has been the practice of a number of economists, from Adam Smith to Irving Fisher, who defined capital in terms of an inventory of man-made resources. For completeness it

would be difficult to improve on Smith's definition of what he terms " stock."

Smith's Definition of Stock

Smith described the stock of an individual as consisting of two parts: " That part which, he expects, is to afford him his revenue "; and " That which supplies his immediate consumption." Examining these two quantities for typical individuals, he arrives at an analysis of " the general stock of any country or society (which) is the same with that of all its inhabitants or members." He analyses stock under three heads:

1. Consumption Reserves—consisting of food, clothing, household furnishings (and presumably any durable consumption goods) and dwelling-houses.
2. Fixed Capital:
   (*a*) Useful machines and instruments of trade.
   (*b*) Profitable buildings which are means of procuring a revenue.
   (*c*) Improvements to land.
   (*d*) Acquired and useful abilities of inhabitants and members of society.
3. Circulating Capital:
   (*a*) Stock of money.
   (*b*) Provisions in hands of producers and retailers.
   (*c*) Raw materials and semi-finished goods. Unsold stocks of finished goods.

This list is of particular interest because it forms the basis of the numerous definitions of capital which have since appeared in economic writing. Different writers have made different selections of items; some have been content with a list of actual assets, and some have preferred to think in terms of the gross money value.

Capital as a Stock of Wealth

Such a list is an inventory of items of wealth existing at a moment of time, and the final content of the list depends on the definition of wealth adopted. It is clear that, unless we are content with a list of physical assets, we must express each item in terms of money value in order to arrive at a summation.

The Status of Titles to Wealth

In making a summation we must avoid double counting, that is, the inclusion of the same item more than once. Smith does this by confining his list to actual objects existing at a moment of time. He does not include titles to wealth, such as securities. We cannot include both the shares of a limited liability company and the real items of fixed and circulating capital it possesses. The National Debt bonds, issued by the Government of a country and held by citizens, represent an asset to the individuals but a debt for the community, and the two cancel out. If, however, we stick to tangible objects and ignore pieces of paper which represent titles to wealth, we shall avoid this difficulty of double counting.

A problem which is very much more important now than it was in Smith's day is that of the difference between the wealth owned by the inhabitants of a country and the wealth existing in the country. Wealth in the country may belong to people living outside, and wealth outside may belong to members of the community. Here we shall probably have no option but to deal in terms of titles to wealth and to add or subtract the net balance.

Human Capital

The list includes, under fixed capital, an item for the acquired and useful abilities of the members of the community. Now it is true that, since the community provides the training, there is a cost to the community involved in the acquisition of these abilities. Also the possession of these abilities by some of its members will generally cause the communal income to be bigger than it would otherwise be. In these respects, acquired abilities resemble improvements to the environment, but where the potentialities of human beings are concerned, our estimates of future income cannot be sufficiently exact to justify using them as a basis of computation. On the ground, then, that any valuation of human beings must be so arbitrary as to be unusable, it is usual to omit this item from computations of capital.

Quantity of Currency as Capital

From the point of view of the individual, a bank deposit or a wad of notes constitutes capital, since it can be readily exchanged for any other form of wealth.

From the point of view of the community, the currency system is a useful instrument. A community with a good currency system is undoubtedly better off than if it had none. The existence of such a system facilitates exchange, enables a more complicated structure of production to be erected, and so enables the income wealth of the community to increase.

This does not mean, however, that the more units of currency it issues the better off a community will be. Provided the currency system is doing its job of effecting exchanges swiftly and smoothly, the size of the issue is a matter of indifference. Why, then, did Smith include the stock of money as part of national capital? His discussion of the matter runs entirely in terms of coin, although the banknote was a perfectly familiar object to him. A stock of gold and silver coin has a value as metal, and is consequently entitled to be included as a commodity stock. Smith appears to have considered the intrinsic value of the coinage was about the same as its face value.

In the case of a paper currency, its intrinsic value is negligible and, although it is a useful piece of mechanism of exchange, no valuation can be put upon it.

Land as a Capital Asset

In Smith's classification of improvements to land as capital, we have the beginning of the classical tradition of distinguishing between the original properties of the soil and improvements made in the land. The considerations which underlay the distinction were probably of an ethical character, and derived from the idea that there is a fundamental difference between exacting a price for the use of assets which are the result of human sweat and toil, and for those which are a free gift of nature, and which must therefore have been appropriated by individuals in a manner not entirely above suspicion. When such an ethical issue exists, a distinction between the two kinds of asset and between the payments to the owners of each is of interest to those who accept and to those who attack the

justification of private ownership of natural assets. When the focus of interest has shifted to the issue of the justification of inheritance of property rights by individual members of the community, the land-capital distinction is no longer of importance on this ground and, since our discussion of rent has shown that there are no peculiarities of land which are significant from the point of view of analysis, there is nothing but the break with the traditions of economic writing to deter us from combining natural assets and produced assets in the one category of capital.

The Inventory Concept of Capital Reconsidered

We have now arrived at a conception of capital as an inventory of all items of wealth which are external to those who own them, existing at a moment of time, and the inventory may include or exclude land according to choice. There is a certain logical completeness about the definition which is attractive. We can, if we like, follow Irving Fisher and construct a parallel concept of income which shall consist of the uses of every type of asset entering into the inventory. When we come to use this concept, however, we shall find it extremely unadaptable. Our capital concept will include quantities of durable consumption goods for which it is virtually impossible to arrive at a statistical valuation, and which do not in any case figure in our economic computations. The quantity of household goods, jewellery, sports and travel equipment, motor-cars and cycles, books and pictures, which a community enjoys makes a substantial contribution to the comfort and grace of living and thus to "welfare," but these items are not readily calculable, nor do they very directly affect other quantities in our economic equations. Therefore, as with the work done by women in their own homes which we excluded from income, we must omit them from a computation of capital. There must be a correlation between our capital and income concepts if both are to be useful, so before examining the definition of capital further let us see what income concepts it must fit.

Income Concepts

As we have seen above, the definition of income which is parallel to the inventory definition of capital is Irving Fisher's

"use" income. Fisher speaks of people enjoying the use of certain goods for a year, irrespective of whether the commodities in question are durable or perishable consumption goods. So income would be composed of the use of so much housing accommodation for a year, the use of so many motorcars, so much furniture and clothing, so many loaves of bread, pounds of tea, bacon and meat, and so forth. The total is summable in terms of money value. For all goods completely consumed within the year, we can take their market price, but for durable items the value of a year's use would have to be calculated, and such a computation would have the disadvantage that it would lack any precise factual basis, although the idea is sound logically.

When discussing the inventory concept of capital, we rejected the idea of including "human capital," but it is clear that personal services must be included in our income concept, which cannot therefore stand four square with the capital concept. The inventory and use approaches to capital and income do not lead to parallel concepts which can form the basis of statistical computations. The concepts are consequently unusable in a method of analysis which is to be applied to real problems.

The use view of income has the further defect that it ignores quantities which are of importance in the demand and supply apparatus which we have built up to explain the basis of payments to factors of production, and we shall therefore examine this apparatus to see whether it suggests other income concepts which may prove more useful. It does indeed provide three possible definitions of income.

First, the object of all economic activity is consumption, so that we may consider income to be the results of the activities we have under review, measured as they pass into consumption. Such a measurement will include both commodities and services. Alternatively, we can view the results of activity as the output of the period in question, and this is not necessarily or indeed normally equal to the consumption of the period. Like the consumption income concept, production income will consist of both goods and services, and can be measured either in physical or money terms. Both will consist of items for which a market price exists. Thirdly, we can concentrate on

the payments made to persons and institutions whose services or whose property are employed in the productive process. This we shall term earnings, and we must now review each of these concepts in turn and explore the relation between them.

Consumption Income

The definition of this concept, used by certain writers, bears a strong resemblance to Fisher's " use " income and suffers from similar defects. It concentrates on the disappearance of goods from existence. This is straightforward enough in the case of single-use goods, but if we are to be logical we must calculate allowances for wear and tear and obsolescence on durable goods, and add them to the value of single-use goods and services used up during the year. Furthermore, if we stick rigidly to the " disappearance " criterion, we must count both the producers' and consumers' goods which disappear, and the risks of counting the same goods and services at both the producers' good and the consumers' goods stages is considerable. Even if we manage to surmount this difficulty, the total we achieve is not very useful in our computations.

It is both simpler and more useful to count goods as they pass into the hands of people we can identify as consumers. We then include both single-use and durable goods at their retail prices. There are few risks of double counting, though the possibility of durable goods which were originally sold to consumers reappearing as producers' goods through the second-hand market certainly exists.

Consumption income, as so defined, is the total expenditure of the period on new consumption goods, or the effective demand for such goods, and it is a useful identifiable quantity which is capable of measurement. No direct measurement of this quantity as consumers' expenditure is possible in Great Britain, but the greatly increased collection of statistics for administrative purposes during the past twenty years has made it possible to estimate annually sales of most consumers' goods and certain consumption services and these have been used in the national income and expenditure computations. The first Census of Distribution taken in 1950 provided a reasonably complete account of the values of sales.

We saw, when we were defining the various categories of commodities in Chapter II, that consumption goods cannot be identified by their nature, but only by the intentions of the purchasers. If we assume, that all people who purchase at retail, purchase for final consumption, and that all purchases for final consumption are made in this way, our definition is reasonably watertight and computable. Consumption income is then the total of goods and personal services purchased by final consumers, and the goods item can be equated to the volume of retail sales.

Production Income

On the other hand, we can regard income from the point of view of goods and services coming into existence in the standard period.

It will now be much more difficult to avoid double counting. One way of doing this is to follow through every commodity and service from the point at which it arises to the point at which it is finally embodied in a finished good.

This is indeed the basis of the method of computation by the Census of Production, which is taken periodically in Great Britain. For every factory and workshop, except the very smallest, a return is made from which it is possible to calculate the " value added " by the manufacturing process. The value of the output of the firm, not merely the goods sold, is calculated at the point where it leaves the factory, and from this figure is subtracted the value of all materials, fuel, power and stores purchased by the firm. By totalling the contribution of all firms and allowing for work carried out by sub-contractors, an estimate, which is free from double counting, of the value added by the process of production can be obtained. To make it complete, we must add the value of the output of agriculture and of mining industries, and of the gas, electricity and transport industries.

The total so obtained is still not identical with the total output of the production process, as we have defined it earlier. Our computation covers this process to the stage where transformation of matter is completed, but the process does not end until the goods are in the hands of the final users. To compute the value added after transformation is complete, we require the

Census of Distribution, which gives the value added by services rendered between the factory gates and the home of the consumer.

It would also be necessary to add a computation of the value of personal services which are directly consumed.

Such, then, are the methods by which a computation of production income could be made, but for the purposes of economic analysis it is useful to look at it in another way, which would be more difficult to follow out statistically, but which brings out one feature of production income which the other concept does not. The measurements given by the two methods should be identical.

In this method we think of the volume of goods which has passed into the hands of consumers, or consumption income, and the change in stocks of materials and semi-finished goods, finished producers' goods and unsold finished consumption goods—over the year.

Thus, wheat will have been turned into flour and flour into other comestibles, which will in turn have passed into consumption income. Iron-ore will have been turned into pig iron, into steel and finally into bicycles, sewing-machines and motor-cars, which will have passed into consumption income or into stocks of finished goods. During the year it is possible to increase or decrease stocks at each stage. If stocks have increased, consumption income has been less in relation to the amount of productive effort put forward than if stocks had been allowed to run down.

Not all of the output of the year will, however, have been used to increase available supplies of finished goods and semi-finished products. Some of it will have been devoted to the replacement of the wear and tear suffered by the fixed equipment of production during the year. Some of this replacement of depreciation must be made if the equipment is to be kept working, even over a short period of time, but some of it can be postponed from one year to another, and it may be decided not to make some replacements at all. In the latter case the equipment will be allowed to wear itself out, presumably because the owners of it consider there are preferable ways of spending proceeds of using the machine. The quantity of

depreciation which businessmen decide to replace is therefore a variable within wide limits, and they will not necessarily decide to restore their equipment at the end of the year to the capacity and degree of efficiency of the beginning of the year.

The community started the year with certain resources of raw materials, semi-finished products, fixed equipment and stocks of finished goods not in the hands of consumers. Let this quantity be R_1 and the similarly determined quantity at the end of the year R_2. During the year a certain output has been produced, and this is the production income we wish to determine. Part of it, C, has passed into the hands of consumers; part of it may have been used to increase stocks of goods of all kinds, that is, to make R_2 greater than R_1, and some of it will have been used to replace the wear and tear of fixed equipment, so that even if R_2 equals R_1 some of the year's output has been used to maintain the stock. We must allow for the possibility that the additions to stock and the replacement of depreciation, D, may not be sufficient to cover the drafts on stock and the year's wear and tear, so that R_1 may be greater than R_2. Then we can define production income as follows:

$$I_p = R_2 - R_1 + C + D$$

that is, production income is equal to consumption income, plus any increase in stocks of all kinds, minus any decrease in stocks, plus any replacement of depreciation.

Earnings

There is yet a third method of measuring income, and this is the total of money incomes of persons and institutions. We term this measurement earnings, though it does not consist exclusively of the type of income the Inland Revenue Department calls earned income. It includes all payments to individuals in respect of labour services rendered, i.e. wages and salaries, but it also includes income from property of all kinds, rents, dividends and interest, and the incomes which " self-employed " people like artists, musicians and bookmakers draw from their activities.

Earnings do not include grants or gifts to individuals " without consideration," such as allowances by parents to children or income from scholarships. These are transfers of income

between individuals, and not payments for service, as are all true earnings.

When we turn to institutions we must be careful again about double counting. We cannot include both the income of public companies and the part of that income such companies distribute as dividends, otherwise we count the dividends twice. As we are concerned not only with the dividends paid by limited liability companies, but also with the sums they put to reserve, the figure for distributable profit is the one in which we are interested. Charities receive income, both in the form of gifts and in the form of income from property. The gifts are part of the income of the donors and must be excluded. Fortunately, the income-tax authorities compute a total of incomes for persons and institutions subject to tax, which covers a very large part of the field of earnings, on approximately the lines we have indicated, though they treat depreciation in a different manner. Additions have to be made for incomes not subject to taxation.

The Three Income Measurements Compared

Let us begin with earnings, for it is the way in which the community decides to dispose of this volume of expenditure that determines what is produced and hence the relative sizes of the other two. Suppose the community were to decide to spend all its earnings on consumption goods, then earnings and consumption income, or the value of the output of consumption goods, would be equal. If it did this, there would be no demand for producers' goods apart from those needed to replace depreciation. If businessmen decided not to replace any depreciation, supposing such a thing were possible, then the expenditure which would have been devoted to that purpose would be distributed as earnings, and production income and earnings would be equal. All productive effort would be devoted to consumption; consumption income, earnings and production income would all be equal.

Normally, however, some provision for depreciation is made, and earnings fall short of production income by that amount. Furthermore, the whole of earnings is not normally spent on consumption; part of them is saved, so that earnings normally exceed consumption income by the amount of savings.

We shall have more to say later on the influences which determine the relationships between these three measurements of income and of the causes of variations in the proportions they bear to each other. Before we can attempt this, there is much to be discovered about what determines willingness of people to consume, their willingness to invest and their willingness to replace worn-out equipment.

Real Income and Money Income

Measurements of production income and consumption income can be made either in terms of actual goods and services or in terms of money value. We can think of an inventory of the goods and services produced or consumed, or of the money value of these inventories. The disadvantage of the inventory is that it is impossible to summarise it in a single figure for comparative purposes. Such a measurement might tell us that in a certain year a country was X thousand tons of steel worse off and Y thousand bushels of wheat better off than it was the year before, but we could not say whether it was better off or worse off on balance.

The use of a money measurement means that we can add up the various items into a homogeneous total, and we can then compare consumption income with production income for the same year and make comparisons with other years. Also earnings can be measured only in terms of money, therefore it is advantageous to make the other measurements in money terms as well.

Money, nevertheless, has a serious disadvantage as a unit of measurement. It is a desirable feature of any good measuring rod that its length shall remain constant. If we are to use money income to make comparisons of wealth at different times, it is desirable that a given quantity of money should correspond to the same quantity of goods and services on both occasions. Unfortunately, this is not the case. An increase in money income need not indicate a corresponding increase in real income; indeed, it need not indicate any real increase at all. Prices may have risen in the interval, so that a smaller quantity of goods may have a higher value on one date than a larger quantity on another date. In making such comparisons, it is therefore necessary to keep an eye on price

changes, and to help us to do this we have the device of the Price Index Number.

Price Index Numbers and Comparisons of Income at Different Times

A very simple form of price index would be obtained by taking the price of wheat per bushel on a certain day each year for a number of years and expressing each price as a percentage of the price in one year, say the first year.[1] If in one year the price of wheat was 50 per cent. above the price in the base year, then we should say that, so far as wheat is concerned, money was worth on the second occasion only two-thirds of what it was worth on the first, and a computation of money income could be interpreted in the light of this knowledge.

Wheat has the advantage of being a commodity which does not change in character from year to year. True, wheat today is not exactly the same as it was five hundred years ago, but the resemblance is probably much greater than for any other article of common consumption which is produced today and which was also produced five hundred years ago. A more important problem arises from the fact that the price of wheat may move in the opposite direction to the prices of other important commodities. It would not then give us a true picture of what was happening to goods in general. To overcome this difficulty, we usually take, not one commodity but a collection of commodities, and price the whole collection on a succession of dates. The bigger we make our collection the more accurately will our index reflect what is happening to the value of money, but at the same time the more difficulty we shall have in finding objects which correspond exactly to each other at different periods of time. The net result is that over short periods of years we can make reasonably good computations of the effects of price changes and so compare incomes, but if we attempt such comparisons over long periods of years, we find that the actual goods which comprise real income have changed so drastically that the things we are comparing are essentially different.

[1] The first attempt known to the writer to estimate changes in the value of money is *Chronicon Preciosum*, Bishop Fleetwood (2nd edition, 1745), which uses the price of corn (wheat) as indicator.

The Definition of Capital Reconsidered

If our capital concept has to be in line with our income concepts, we must concern ourselves only with those items which are included in our economic accounting. Thus, all consumption goods pass out of the account as soon as they are purchased by final consumers, so that durable consumption goods will not constitute part of capital once they are in the hands of consumers. We can, if we like, create a further capital account for them and label it consumption capital, for such a total is useful in any attempt to estimate the welfare which people derive from the process of production and consumption. The nature and purpose of this stock is so different from that of capital, considered as a factor of production, that the two must be kept distinct.

If it is objected that the two groups of items may be difficult to distinguish, since their classification depends on the owners' intentions, it may be pointed out that the intention is clear from the type of account out of which payment is made. Private individuals have private accounts into which they pay their earnings and out of which they finance both consumption and savings. Both individuals in their capacities as entrepreneurs, and also firms have business accounts from which they pay the expenses of production and into which they pay the proceeds of production. The process of increasing the quantity of capital is termed investment, and consists of buying producers' goods and paying for them from business accounts. We shall think of investing as the process of acquiring assets which consist of producers' goods of all kinds; the total of such assets existing at a given time is capital. The category of capital is therefore confined to producers' goods, and it belongs to the owners of business accounts.

Now capital is commonly regarded as being accumulated by saving, so we must examine the nature of this process. We have seen that earnings consist of the payments made to the factors of production for the part they play in the productive process. The total of these payments is equal to the total value of the output of the period, less any wear and tear of fixed equipment which is replaced, and they are due to the owners of private accounts into which earnings are paid and are then

at the owners' disposal for expenditure on consumption or for saving. In the case of firms, however, some earnings are commonly retained in the business account, and are not made over to the individuals who really own them. Savings are paid from private accounts into business accounts as loans, or are retained there by the process we have just described, in which case the final effect is the same as if the process of payment into a private account, saving, and return to a business account as a loan, had been followed. Capital is therefore created by investment, which consists of the purchase of producers' goods paid for from business accounts, while saving is a process of refraining from expenditure on consumption, and the income so released may be used to make loans to increase the funds available in business accounts for making purchases of producers' goods.

Capital as a Factor of Production

We have defined the capital of a community as the quantity of producers' goods it possesses, that is, its stock of goods of all kinds required to make other goods: raw materials, partly finished goods, goods awaiting sale, and equipment of all kinds. Now we can talk of the productivity of capital goods: the contribution they make to income.

When capital goods are used with land and labour, the output is increased in two ways: the productivities of land and labour are themselves increased and capital goods have a marginal product.

Suppose we have two farms equal in size and in every way identical; the same number of men is employed on each, but the quantity of capital employed on the first is greater than that employed on the second. Then the increase in product obtained by employing one more labourer on the first farm will be greater than the increase from employing one more on the second. Labour is more productive when given capital to assist it. Similarly, if we keep the quantity of labour constant and increase the quantity of land, the same condition will hold.

If, on the other hand, the quantities of labour and land remain the same and the quantity of capital is increased, we shall observe the usual " Diminishing Returns Law " effect.

Under normal conditions, where the point of maximum marginal returns will have been passed, the addition of a further unit to the community's stock of a particular producers' good will increase the output of some products, but by a smaller increment than that due to the addition of the previous unit. So an increase in producers' goods as a whole, at a given point of time, will increase the income of the community measured in physical terms, but as capital is continually increased, the rate of increase of income will diminish.

So far then, capital appears to behave like the other factors of production, but whereas land and labour are commonly hired and we compute the charge for hiring them, capital goods are bought and sold, and we must calculate the price or capital value.

To its owner a producers' good represents a prospective stream of income stretching away into the future. Its capital value is the value of the income which it is expected to produce before it is worn out, and this may be greater or less than the cost of producing it. Thus, at the present time, the market price of houses, i.e. the price people are willing to pay for the satisfaction of living in them, which constitutes the annual income, is greatly in excess of the current cost of construction. On the other hand, in the depression years of the 1930's it was possible to buy ships or textile looms, which nobody could employ, for a small fraction of the current cost of construction less depreciation. In the first case it will pay to construct more capital goods, so long as the state of affairs in which they can be sold for more than their construction cost continues to exist, but the increasing stock of such goods should eventually bring about a fall in the market price until capital value, in the sense of the present value of the expected income, is equal to the construction cost. Similarly, if capital value falls short of construction cost, no one will be willing to undertake construction, so that, through wear and tear, the stock will eventually diminish and the capital value of those left will rise to equality with construction cost.

If the stock of a given type of capital good is therefore to be in equilibrium, showing no tendency to either increase or diminish, construction cost must be equal to the present value of the expected income stream.

The Pricing of Capital Goods

The value of a capital good is thus the "present worth" of a series of net receipts which will be derived from it during its lifetime. If for simplicity we assume that there is one receipt of £A each year for T years, the value of the asset producing this income stream is the present worth of £A per annum for T years when the current rate of interest is r. Then, as any school algebra will show, this present worth is equal to

$$\frac{A}{R-1}\cdot\left[1-\left(\frac{1}{R}\right)^T\right]$$

where R stands for $1+\frac{r}{100}$. The capital sum so obtained is the value of the asset producing the given income. It will be greater the longer the period of time, and will also vary according to the rate of interest. If the rate of interest falls, the sum necessary to buy a given income will rise, and a rise in the rate of interest will reduce the capital sum.

This rule will apply to the value of any asset promising a certain income per year for a stated time. If the income goes on for ever, as with a perpetual ground rent or an unredeemable security such as "Consols," the formula is somewhat simpler, but otherwise the effect is the same. An alternative method of stating the price is at so many years' purchase, that is, the asset is worth a stated number of years' income.

When the capital asset is a machine or a block of offices, the estimation of the income per year which will be obtained from it is more uncertain. Estimates will, however, be made. It will be expected that a given machine, if installed, will save so much a year in wages, and as it will cost so much each year for maintenance, the machine will "pay for itself" in so many years. Similarly, the purchaser of an office block will calculate that, after paying costs of upkeep, the income from rents is likely to bring in a certain sum, and that he will get his money back in a certain number of years.

This does not mean, of course, that the purchaser hopes merely to get his money back before his asset is worn out or fit for demolition. He is making the investment to obtain an income, and this is merely his way of comparing the profit-

ability of different opportunities to invest which are open to him. He might use our geometrical progression formula to calculate the rate of return, knowing the amount of investment required and estimating the probable annual income and the length of time it is likely to continue. He happens to find the " number of years' purchase " method simpler, but the effect is just the same.

The value of a durable capital good is therefore the anticipated annual income capitalised at a rate of interest.

What that rate may be and how it will vary for different kinds of assets we must leave for later consideration. The important point here is that the capital value of an asset depends on the income which can be expected to accrue to the owner of it. Income, then, is the fundamental concept, not capital, and, knowing the income derived from an asset, we can discover the capital value with the help of a rate of interest.

Interest Rate and the Value of Capital as a Whole

An important point which has emerged from this discussion of the pricing of capital goods is that the quantity of capital, i.e. the monetary fund invested in producers' goods, can vary quite independently of the income the goods are anticipated to produce, owing to changes in the rate of interest.

This fact makes comparisons of the value of the capital possessed by a community at two different times a difficult matter. First we must start with the incomes from capital at the times in question. To sum these we must use money prices, and thus become involved in changes in the price level. Then, even if we should happen to satisfy ourselves that the income streams at the times in question were equal, our calculation might show an increase or decrease in capital according to whether the rate of interest had fallen or risen.

It is quite possible, therefore, that making a computation of the value of the capital of a community on two different occasions, we might arrive at the same figure on both occasions, and yet it might be clear from evidence of destruction of capital in the interval that in fact the community possessed fewer assets on the later occasion. This could come about if, on the first date, income had been bigger and the rate of interest higher, and on the second, income and the rate of interest had

both been lower. It may be argued that this case is not a very probable one, that a destruction of assets would lead to a rise in the rate of interest, not to a fall; but the case is not impossible, and it does bring out the unreliability of comparisons of measurements of communal capital taken at different dates.

Neither can we get any help in making such comparisons by abandoning the money measure and attacking the problem in terms of real capital goods. A comparison of complete inventories, taken at two points of time, will not tell us whether capital has increased or decreased in the interval, if the kinds of items included have changed. Because of technical progress such changes are very probable. An item in one inventory may bear the same name as an item in the earlier inventory, but they may have very different powers of producing income. We cannot allow for this change in capacity by adjusting price and by counting as replacement the cost of an asset capable of producing the old rate of output, because when an asset becomes very obsolete before it is scrapped, it may be more costly to replace it by a replica than by the more productive modern version. A 1920 model car would cost more to produce now that all moulds and patterns are destroyed than its more efficient modern equivalent.

It is clear that we cannot talk of the total quantity of capital in the same way that we can talk of the total quantity of land or the total quantity of labour. These latter are totals which exist irrespective of the income that is being derived from them, but we can measure capital value only if we first know the income derived from capital assets and the current rate of interest. We do not need to know wages before we can know the quantity of labour, or rents before we know the quantity of land, though we do need to know both rent and interest rate before we know the value of land.

We can find such a relationship between quantities of particular capital goods and the sums it would be worth while paying to hire them, and in this respect there is no difference between capital goods and land, except that the quantity of capital goods can be voluntarily increased. We can talk of the marginal productivity of particular types of capital goods, but not of the marginal productivity of capital as a whole, because we cannot measure capital as a whole unless we first know its product.

Fortunately, a measurement of total capital is not essential to the solution of any problem we have to face. This has not always been the case. The classical English School of Economists required to know the "Wages Fund" or quantity of capital before they could determine the rate of interest, and the productivity theories of interest determined that rate by a comparison of the quantity of capital and the marginal product of capital. We do not need it for this purpose, and though, as we shall see, the quantity of capital goods in existence influences the rate of interest, yet we do not need to know the total value of them in order to determine the interest rate.

Further Reading

Fisher, Irving: *Nature of Capital and Income*, Part I.
Hicks, J. R.: *Social Framework*, Part 3.
Pigou, A. C.: *Income.*

More Advanced Reading

Keynes, J. M.: *General Theory of Employment Interest and Money* Chaps. 4–6, 16.
Keynes, J. M.: *Treatise on Money*, Vol. 2, Chaps. 27, 28 and 30
Pigou, A. C.: *Economics of Welfare*, Chap. 4, 3rd Ed.

Chapter XVI

INCOME, SAVING AND INVESTMENT

WE must now take up the further examination of the three measurements of income which we defined in the previous chapter. There we saw that income could be measured, first as the value of the quantity of goods and services produced each year; secondly, as the value of goods and services passing into the hands of final consumers, or consumption; and thirdly, as earnings, or the total of personal incomes of individuals or institutions. We saw further that production income normally exceeds earnings by the amount of depreciation replaced, and earnings in turn exceed consumption income by the amount of investment, so that we have the following relationship:

Production income = Earnings + Depreciation replaced.
= Consumption income + Investment + Depreciation.

Of these three concepts of income, the really vital one for our purposes is earnings, for it represents the demand for goods and services over the income period. It represents the total sums transferred from business accounts to private accounts, or, to be more accurate, the sums transferable from one kind of account to the other. It is therefore the total of sums spent on consumption or handed back to business accounts as savings.

Depreciation

Depreciation is the total of replacement of wear and tear which businessmen choose to make. Some writers have claimed that depreciation should be computed on such a scale as would restore the total quantity of capital to the level of the beginning of the year, but as we have seen, we can attach no meaning to the total value of capital, which changes with

changes in the rate of interest. If the rate of interest rises, the value of all assets yielding a fixed income falls, but no one would suggest that businessmen should take such a change into account when deciding the amount of income which can be distributed as earnings. Similarly, if a fall in the rate of interest caused sufficient increase in the value of assets, it might be that the value of assets at the beginning and end of the year was the same, although no physical replacement of wear and tear had taken place.

If, on the other hand, we think of replacement in physical terms, we are faced with the problem of obsolescence. If technical change is taking place, worn-out pieces of equipment will not be replaced by replicas, which might indeed be much more expensive to procure than superior current models.

A certain amount of replacement must take place, if production is to continue even in the short run. How much depreciation actually is replaced will depend on the view businessmen take of the future profitability of using that equipment and on their estimates of the level of costs of replacements at the moment compared with the prices expected to rule at some future time. If the firm considers its requirement for a certain type of equipment in the future is likely to be less than in the past, it will not replace all the depreciation arising from the year's working, and earnings will, to that extent, be swollen by sums which represent a disinvestment of capital.

From the point of view of the individual firm, its production is the value of its net output; that is, the value of goods sold or available for sale less the expenditures incurred to produce those goods. A certain provision will be made to restore, at least partially, the damage to equipment caused by the productive process. The balance of net output is earnings.

Saving and Investment are Equal

As we have seen earlier, we can look at earnings in three ways. We can regard it as the total payments to factors of production for the services performed in the production of consumption and investment goods; it is the total spent during the same period on the two classes of goods; it is the total expenditure on consumption and savings. That is:

Earnings = Cost of consumption goods + Cost of investment goods.
= Expenditure on consumption goods + Expenditure on investment goods.
= Consumption + Savings.

These three equations must be identities. What is spent on consumption goods is equal to what is paid out to all factors of production for making them. For the case of perfect competition this is obviously true, as total receipts equal total costs at market prices of the factors employed. Under monopoly, if we include the monopolist as a factor, total receipts are equal to total distributed incomes. The same is true of costs of production of investment goods and receipts from their sale. So, if the item consumption is the same in all three equations and investment is the same in the first and second, it must also be the same in the third equation. Expenditure on investment goods or investment must therefore be equal to savings.

Saving and Investment are Simultaneous Processes

The necessary equality of these two quantities at all times, and not merely over periods of time, is apparent when it is realised that saving as well as investment involves a transaction between two parties. In order to save it is necessary to acquire an asset, either an investment good which has just been produced, a security or cash. In any case, someone must part with something, so that an act of saving from the individual point of view can take place. If one individual saves by spending part of his income on a new investment good, then from the communal point of view equal increments have been made to both saving and investment. The purchase of an old security, or the accumulation of cash, merely involves a transfer between two members of the community, and it is necessary to go a step farther and discover what the person who parted with the cash or the security does, before deciding that any act of saving or investment has taken place. In the case of saving taking the form of a subscription to the capital of a new company, it is necessary to enquire whether the purpose of the issue is the transfer of ownership of an old investment, or the

acquisition and organisation of new investment goods into a business. In this respect what is true of a Crusoe is true of a community. A community can save only by using resources to make goods which will produce goods in the future, instead of using them to make goods for current consumption.

From this two highly important conclusions follow. One we have already arrived at—savings must equal investment. The second follows from the first—the processes of saving and investing must go on simultaneously. There can be no question of a community saving first and then investing. An individual can do this, but not the community.

Savings and investment so defined must therefore be identically equal in all circumstances, and there can be no question of them being equal at a price, that is, a rate of interest, as in the case of supply and demand for a commodity.

Yet this equality of savings and investment is the result of two different sets of decisions—a set of decisions to save and a set of decisions to invest. If all people saved and invested their own savings, the matter would be simplified, but there are many who save and do not invest, i.e. do not buy producers' goods, and there are others who invest but do not save. The latter group borrow the savings of the first. The further explanation of the causes of this equality must, however, be postponed for the moment while we explore the basis of decisions to save and to invest.

Saving is Abstention from Consumption

Saving, whether for the individual or the community, is the difference between earnings and consumption. Consequently, we can consider the problem from two opposite angles, from that of willingness to consume or from that of willingness to abstain from consumption. It was willingness to abstain from consumption which first received attention from economists, and Senior defined a third factor of production to rank alongside land and labour which he called " Abstinence." Capital is created through saving by those who abstain from consumption. To induce them to undertake the unpleasant task of reducing their consumption, some reward has to be offered to those who abstain, and so Senior arrived at his theory of interest.

Later writers developed this idea, but instead of envisaging savers as *forgoing* consumption, thought of them as postponing consumption. So we got a succession of "agio," or Time Preference theories, which pictured an individual as transferring consumption from the present, decreasing consumption income now and increasing it at some future time.

In order to do this the individual has the choice of collecting a stock of consumption goods which will actually be consumed in the future; of collecting a store of money which it is intended to spend on consumption goods in the future; of buying an asset which will directly produce income for consumption in the future; and finally of lending money on terms which provide for periodical payments of interest.

The first of these is a favourite of many of the early writers on capital and interest; there is Roscher's fisherman who saved up salted fish until he had enough to live on while he built a boat and made a net. The second was the method of the rich merchant of the seventeenth century who, when he retired from business, took with him a stock of golden guineas on which he proposed to live for the rest of his life. It has been the traditional method of saving of peasants the world over, until the present time. The third method is that of purchasing a house to provide a real income or some productive asset which will provide an increment to money income. The fourth method redistributes money income through time.

Individual Saving and Communal Saving

If an individual accomplishes any one of these actions, then from his own point of view he has saved, but from the communal view-point this is not necessarily so. In the first case, where consumption of actual goods is postponed, there has been no addition to capital as we redefined it in the last chapter. The hoarding of money merely transfers the stock of money existing in the community from one set of hands to another. In the fourth case we have to wait and see what the borrower does with the loan before we can tell whether there has been any addition to the assets of the community. From the communal point of view, the only acts of saving are therefore those which result in the production of new assets, i.e. which are at the same time investments.

When we are dealing with the time preference view of saving, we are dealing with the sums individuals would like to save, and the total of these sums has no necessary connection with what they actually do save as a community. We cannot speak of what people would like to save in the abstract, we must relate it to the incomes from which the saving must take place, and this presumably is the income which people expect to enjoy in the period for which they are planning. They are looking ahead, three months, six months or a year, and during that period they expect their incomes to be so much, and out of those incomes they intend to save so much.

Willingness to Consume

The second method of thinking first of the part of income which is consumed and then of savings as the unconsumed balance is that of Lord Keynes, who discusses the propensity of people to consume a certain proportion of their income. The proportion of income consumed decreases with an increase in income, so that although people do consume more if income is larger than if it is smaller, the proportionate increase in consumption is less than the proportionate increase in income. Similarly, when income falls people reduce their consumption, but not in proportion to the fall in income.

This proposition is probably more accurate when expressed in terms of real income rather than of money income. People have an accustomed standard of consumption which they endeavour to maintain. If the prices of commodities rise at the same time that money income increases, and it so happens that the higher money income will buy the same collection of goods and services as before, but at higher prices, then there would be no reason to suppose that the proportion of income consumed would change, whether measured in money or in other terms. If changes in commodity prices and in money income do not compensate, the proportion of income consumed will be different according to whether the computation is made in terms of money or in "real" terms. To avoid this difficulty, Keynes measured both income and consumption in terms of a wage unit, i.e. the price of a standard quantity of labour. An alternative would be to use consumption units or standard parcels of goods.

Does the Rate of Interest affect Savings?

The exponents of the Abstinence and Time Preference theories of interest considered that the quantity of savings varies with the rate of interest, and that change in the rate of interest was the chief cause of such variations. They admitted that there would be savings even if the rate of interest were zero, but they did not explain why some people always save in such a way as to earn no interest even when they could obtain an income from their savings. Yet they were willing to draw an upward sloping supply curve for savings, showing savings increasing with the rate of interest.

If we consider the probable behaviour of various types of saver, the justification of such a curve appears extremely doubtful. Savers may be grouped into three broad classes:

1. The small " rainy-day " savers.
2. The savers who wish to provide for old age, retirement or for dependants after the death of the saver.
3. The " automatic " savers, corporations and the very rich.

The first of these groups are savers for emergency, and such motives influence the majority of savers in some degree. Even the very rich will find it convenient to have funds in some form where they are readily accessible to cover unexpected expenditure. The savings of the small man will consist entirely of emergency reserves, and the sums involved will be too small for any interest received to make a significant addition to income. In all cases of such reserves it is the capital sum available which is important, and the rate of interest will not affect the amount in a significant degree.

The second group are saving with a future interest income in view, and it appears equally arguable that on the one hand they will be encouraged by a higher rate of interest to secure a bigger future income, or that they will take advantage of a high rate of interest to secure the desired income with less saving.

In the third group the automatic savings of rich men result from the tendency to maintain a constant standard of consumption. The balance of income is saved. Corporations, too, tend to maintain a customary level of dividends and to place the balance to reserve; that is, they enforce savings on their shareholders.

It is the last of these groups which, in fact, will provide the bulk of saving. In the first group the rainy days occur against which provision is being made, and the total saving probably remains tolerably constant, or at least stands in a fairly constant relation to income. Similarly with the old-age type of saving. Young people are adding to the pool and old people are withdrawing from it. The pool will grow with growth in the income of the community, and again bears a fairly constant relation to income. The third item, we have seen, varies with income.

The result, then, of this attempt to connect the amount of saving with the rate of interest has been to confirm the view that communal savings depend primarily on communal income. We can, if we wish, express savings as a function of income, but as savings are equal to the difference between income and consumption, and we require the relationship between income and consumption for another purpose, it is more convenient to concentrate on the portion of income which is consumed rather than on the balance which is saved.

Investment Opportunities

In the previous chapter we did some violence to popular usage, and defined an act of investment as the purchase of an asset which is expected to yield either a service or a product in the future. Thus, in our terminology the purchase of a security is not an investment; it is one way of making a loan. The true investor is the seller of the security, or the borrower, provided he uses the proceeds for the acquisition of assets, and, in general, securities are issued and loans made with a clear understanding as to what is to be done with the proceeds.

An asset is purchased because it is believed that over a period of time it will produce a certain income, and the prospects of that income are considered to make it worth-while to incur the cost of the investment.

There will at any time be numbers of opportunities to use a capital asset for the purpose of deriving an income. There will be possibilities of using capstan lathes for making fuse mechanisms, or parts of vacuum-cleaners, or cigarette-lighters; building materials may be used to build houses, or cinemas, or factories. The Government might decide to use the output of electrical

apparatus to electrify the main-line railways of the country or to devote a large part of the steel output to building road bridges over the various estuaries round our coasts or to divert materials from the motor industry to the supply of air-conditioning plant for all houses.

All sorts of projects are technically possible, but the value of the services or products each of them can produce might represent a wide variety of rates of return on the cost of the investment. Thus, an investment of £1 million in cinemas might bring in a return of £100,000 a year or 10 per cent., while if invested in the electrification of railways it might give a return of 5 per cent., and when invested in a road bridge a return of only 2 per cent. The percentage which the expected income bears to the cost of the asset is termed by some writers the " Rate of Return over Cost " and by others the " Marginal Efficiency of Capital."

Furthermore, although the investment of £1 million in cinemas might bring in 10 per cent., the return to a second million might be something less. There would be diminishing returns to investment. Similarly, while the electrification of a busy railway route might earn 5 per cent., the next choice might bring a smaller return, and there might be some small branch lines where the return to the investment would be negligible.

So it would be possible to arrange investment opportunities in a descending order of rate of return over cost. Clearly it will be possible to find a larger number of opportunities of investing capital to produce 10 per cent. than to produce 20 per cent., and many more which can produce 1 per cent. than can produce 10 per cent. The schedule of rates of return over cost will therefore show smaller quantities of capital for high rates of return and large quantities of capital showing low rates of return.

We shall therefore have an inverse relationship between rate of return over cost and quantity of new investment, which when plotted will give a curve that slopes downward from left to right like a demand curve. It is indeed very like the curve of demand for capital based on the marginal productivity of different quantities of producers' goods as used by many writers. We have ourselves used a demand curve of this type for labour and land. Why do we not do the same for capital?

Marginal rates of return over cost or marginal efficiencies of capital depend on the marginal productivities of producers' goods, but are not identical with them. Marginal productivity is a measurement of the short period, ideally a measurement taken at the present instant of time. Its value depends on the particular diminishing returns law which is true at the moment. The marginal efficiency of capital *now* depends, not only on the marginal productivity of capital goods now, but also on what the marginal productivities of these same goods will be at every stage during their life until they are worn out or obsolete. Investment is undertaken, not merely in the light of existing conditions, but in the light of all the prospects of return from it.

Furthermore, we considered the return to labour and capital as certain, though this is not strictly accurate. The results of processes which are taking place at the present moment are not exactly predictable; outputs may be more or less than is expected. When, as with the marginal efficiency of capital, we are concerned with estimates of what marginal productivities will be over a period of years, the outcome is still less certain. The complications arising from this cause we shall deal with in the next chapter.

Looking Forward and Backward

We must now make some comparisons between these accounts of willingness to save and to invest.

The Time Preference Theory of Saving attempts to forecast savings. It tells what savings people would like to make, but neglects any possibility that the effects of one person's endeavours to save may defeat the intentions of another. It is the total of what separate individuals would like to save if their incomes should reach the expected level. If incomes for some reason do not reach this level, then savings also are going to be less than was expected.

If we think of saving as what is left over after willingness to consume has done its work, we may be either forecasting or recording a result. If we say we expect income to be so much during the next year, then propensity to consume will tell us how it will be divided between consumption and saving.

The schedule of marginal efficiency of capital is also a forecast; we stated it in terms of expected returns. The forecasts

which form the basis of willingness to save and willingness to invest are made on different grounds and by different people, so there is no reason to expect that the amount people will wish to save in a given set of circumstances will equal the sums people wish to invest in those circumstances.

These concepts of saving and investment have been termed *ex ante* saving and *ex ante* investment respectively. As we have defined *ex ante* saving, it can be represented as a schedule of quantities of saving which people will be willing to make at each of a scale of values of expected income. The time preference theories also made *ex ante* estimates of what savings would be for each item of a scale of interest rates. *Ex ante* investment is again the schedule of quantities of investment people will be willing to make at each of a scale of expected marginal efficiencies of capital. There is therefore no necessary equality between savings and investment viewed *ex ante.*

If, however, we look back on a period of time, we know that the sums which have been invested in that period must be equal to the sums which have been saved in that period. This comes about because, although people may *attempt* to save by acquiring cash or old assets in the possession of other people, the only *effective* acts of saving are the acquisition of new assets; that is, acts of investment.

These are *ex post* saving and *ex post* investment, and they are not schedules but single quantities. They are the records of what actually took place during the period under review, and are necessarily equal.

Savings and Interest

This discussion of savings and investment has given us no clue as to the nature of interest, apart from the negative conclusion that it is not the reward for saving. We have failed to find any causal relationship between interest and saving such as we have found elsewhere between supply price and quantity supplied.

One possible line of investigation did, however, show itself. The savers and the investors are not necessarily the same people, and in order that the savings of one person may be available for investment by another, it is necessary that the saver should make a loan to the investor. Savings are earnings

which are not spent on consumption, so they accrue in the first place in the form of money. We may therefore find the answer to the question, "For what is interest a payment?" in the terms on which people are willing to make loans. Before we can do this, we must first investigate some of the properties of money.

MORE ADVANCED READING

HICKS, J. R.: *Value and Capital*, Chap. 14.
KEYNES, J. M.: *General Theory*, Chaps. 7–12.
MYRDAL, G.: *Monetary Equilibrium*, Chaps. 1–5.

Chapter XVII

INVESTMENT, RISK-BEARING AND PROFIT

We must now develop the picture of the process of investment which we sketched roughly in the previous chapter. Investment consists of the acquisition of producers' goods which it is believed will produce during their lifetime an income, the present value of which is not less than the cost of the investment at the time it is made. The investment may take the form of purchase by a merchant of a stock of a commodity which he hopes to sell for a sum greater than its cost to him, plus all the expenses for purchase, handling and resale involved. Alternatively, it may take the form of purchase of materials by a manufacturer who expects, from his activities, a surplus over and above the expenses of all kinds which he incurs, plus a remuneration for his own personal activities. Again, the investment may be in a piece of durable equipment, or a building which is expected to yield during its lifetime a series of instalments of income which in total will exceed the original payment for the asset.

Investment is undertaken, therefore, in expectation of a certain quantity of net receipts, whether these accrue as a single sum as in our first and second instances, or as a series of instalments over time as in the third. The explanation of the amount of this net receipt and the relation between that amount and the size of the investment has traditionally been termed the theory of profit, but although writers on this subject have been fairly well agreed on the quantity they were explaining, they have differed considerably in their views of the nature of the payment and the identity of the recipient. The differences have been attributable to differences in the organisation of industry in the economic systems with which the various writers were acquainted.

Theories of Profit

The English Classical School of writers originally identified these net receipts with the " profits of stock," and regarded them as the return to the owner for the use of his capital, but one by one certain elements were eliminated from it. It was pointed out that the profits of stock so defined commonly included the wages of superintendence of the owner-manager of the business and also a payment of interest on the capital employed, because in the English nineteenth-century business world management and capital were usually provided by the same persons. Still, from this attempt to explain profit we derive one relevant idea. Profit is a contingent reward and not a contractual one. Its amount depends on the success of the operations from which it is derived.

A further attempt to explain profit concentrated on the function of the entrepreneur, the individual who sees an investment opportunity and organises factors of production to take advantage of it. A further examination of the functions of the entrepreneur showed flaws in this theory. In a world where there was no change, where economic processes repeated themselves year after year, the problems of initiation and co-ordination of production, which are the business of the entrepreneur, would not arise. There would be no particular advantages in the entrepreneur form of organisation, and producers' co-operatives and consumers' co-operatives would be easy to organise and equally efficient in all lines of production, instead of, as now, being confined to certain fields of activity.

The functions of the entrepreneur are therefore concerned with the uncertainty of the future, so that it appears that profit has something to do with the risks and uncertainties which arise from the fact that the future is unknown. A further difficulty arises, however, in that it is not easy to identify the entrepreneur in the contemporary organisation of industry and commerce. The initiation of new enterprise is commonly undertaken by a financial house, which specialises in floating new companies with the object of selling them to the public if they are successful. It does not expect every venture to be successful, but looks to make sufficient profit on some of them

to cover the cost of all its operations and leave an adequate margin of net receipts. The actual conduct of production is, under contemporary conditions, largely organised by paid managers who may have no financial interest in the concerns they manage, and who earn a salary which is connected only in the broadest fashion with the success of the concern.

Investments and Loans

At this point let us return to the distinction we have made between investment and the making of loans. If we think of investments as purchases of physical assets, i.e. of producers' goods, and of all transfers of funds from private accounts to business accounts, which are not payments for goods and services, as loans, we shall be separating two quite dissimilar functions and achieving a clearer view of a matter which has long presented a confused appearance. We shall regard all those titles to wealth, or securities, which are popularly termed investments, as acknowledgments of loans. The private citizen who buys shares in an industrial concern may be "making an investment" when looking at the matter from his own point of view, since he has provided himself with a potential source of income, but, from the point of view of the community, all that has happened is that ownership of a loan has changed hands.

Our study of the relationship between the amount of investment and the size of the stream of income which results from it, we shall term the theory of profit, and the relationship between the size of a loan and the annual payment made for it, we shall term the theory of interest. In the present chapter we shall be concerned with the first of these problems.

Risks of Investment

The conduct of any business activity involves investment which is attended by risks of events happening which will make the results of the undertaking differ from what was expected when the investment was made. We commonly associate the term risk with the possibility of the occurrence of an unpleasant event, but here we are really concerned with the possibilities of the results of investment differing from the expected result in either direction. All forms of property are

subject to risks of " fire, flood and tempest, acts of God, King's enemies and civil insurrection " as the insurance policies say. Many kinds of plant are subject to explosion risk; the physical results of no industrial process are absolutely certain, and in addition there are the economic risks: that demand will differ from that which is forecast; that technical change will render the methods of production obsolete and possibly cause the producers' goods in which investment has been made to be worthless; that price changes of factors, materials or products will alter the financial results of the operation from that expected. Not all these risks are necessarily associated with profit, and we must classify them and eliminate those which are not. It is possible for an enterprise to reduce some of these risks to negligible proportions, and to replace a possible source of loss by the certainty of a known cost.

If in a competitive market some form of production were completely free from risks (that is, the results of investment were absolutely certain), the total receipts from the sale of output would be just sufficient to pay all factors of production their market prices without any surplus. Each producers' good used in the process would, during its lifetime, produce a sum sufficient to repay the loan which was raised to purchase it, together with interest at the rate for perfectly secure loans ruling in the market. In a world in which there are risks, firms will buy producers' goods if they consider that the income from them will suffice to pay off the loan they had to raise; to pay the higher rate of interest lenders will demand because the loan is risky; and, above that, to leave a surplus which makes it worth-while running the risk that the income will not in fact suffice to allow them to pay their obligations.

Classification of Risks

There are three general methods of dealing with risks, and these afford a convenient basis of classification:

(1) Risks may be cancelled by offsetting opposite risks. Sometimes it is possible for the risks themselves to cancel out, or, if the number of individual risks to be borne is large enough, it may be possible to calculate the mathematical probability of the event whose occurrence is feared.

(2) Some risks, although not exactly calculable, may be such

that persons with special skill in, or liking for, risk-bearing may be willing to assume them.

(3) There remain the residual risks, which must be borne by those who conduct the enterprise.

Each of these three classes may be further subdivided, and we will now consider some of the chief methods of providing for risks in common use.

Elimination by Organisation

Sometimes it is possible to arrange the organisation of a firm so that risks are eliminated. In such cases the risk can generally be estimated fairly accurately, and it becomes a question whether the cost of the reorganisation is more or less than the probable loss, if no effort is made to eliminate the risk. In a shop there is the risk of theft of goods by customers. It has been said that in certain London department stores such thefts amount to 5 per cent. of turnover. This loss could be greatly reduced if the freedom of customers to examine goods were restricted; but it is considered worth-while to incur the loss through theft rather than to suffer the fall in turnover which would follow from restriction of display.

When considering degrees of division of labour, we encounter cases where risks attaching to production, like weather risks in farming or risks of changes in demand, lead to an organisation of production which, although not the most efficient from the technical point of view, is nevertheless the cheapest in the long run, because it is readily adaptable to meet changes in circumstances.

Sometimes there are economies of size to be found. From the law of large numbers, we know that the larger the number of cases liable to a certain occurrence we take, the more accurately can the number of cases in which the event actually occurs be calculated. Thus, in bottling champagne a certain number of bursts occurs. The proportion of burst bottles can be estimated from past experience. The larger the bottling plant, the more closely does the realised number conform to the probable number. Where the occurrence of an event can be calculated precisely or estimated empirically, it ceases to be a risk and becomes a cost. The loss on the bottles of champagne which burst is part of the cost of production of the bottles which do not burst.

An element of risk remains, however. It is, that in a particular enterprise the number of bursts will exceed the " probability." If a penny is tossed, the chance of it coming down " heads " is one in two, or a half. This is termed the " probability " of the event of a head turning up. If a dozen tossings were made, the number of heads might be more or less than six; it might be nil, or it might be twelve. If we did a thousand sets of twelve tossings, some cases of no heads and others of twelve heads might turn up, but cases of six heads would be in the majority. As we increased the number of tossings in each set, the more closely would the results conform to the theoretical probability of one-half heads. If we did a thousand sets of a thousand tossings, the majority of sets would differ little from five hundred heads. The case of a thousand heads is not impossible, but the chance of it occurring is very small. The risk involved in such cases is that the realised result will differ widely from the expected result; the chance of this becomes smaller and smaller as the scale of operations is increased.

Insurable Risks

If a risk is calculable there is the possibility that insurance against it may be organised. But before a risk becomes insurable, it must fulfil the additional condition that insurance of it will not greatly increase the probability of its occurrence. Ideally, the insuring of a risk should not make the occurrence of the event more frequent, but of all the risks insured against the Jupiter Pluvius policy alone conforms to this standard. Life insurance conforms approximately; for though rare, it is not unknown for an individual to commit suicide in order that someone may benefit from insurance policies, and criminal records provide other evidence. Insurance against fire, sickness, unemployment and motor-car accidents all require stringent safeguards to prevent fraud.

The insurance company operates through an application of the law of large numbers. The period of life before any one man of, say, thirty-five is quite incalculable, the risk of hardship for his dependants in the event of his early death is great, but the proportion of all men of thirty-five who will live to be thirty-six, or forty, or seventy, or any other age can be estimated quite accurately from the available data. An insurance

company uses the data of the mortality tables to calculate rates at which provision for an early demise can be made. The risk to individuals has been turned into a virtual certainty for the company that it will have to pay certain sums per annum on its policies. The risk for the company is that it will not get enough individual risks for the law of large numbers to operate. Consequently, its payments may exceed the " probable " amount. Insurance can be applied to any field of risk where there is available data which can be analysed with the help of the theory of probability and where the introduction of insurance will not greatly alter the nature of that data.

Transfer to Specialists in Riskbearing

Some risks cannot be reduced to an actuarial computation, and so are not insurable; but it may, nevertheless, be possible to arrive at a subjective estimate of the degree of risk involved. Provided that the circumstances in which the risk occurs are outside the control of the person running the risk, we have a field for the professional risktaker and possibly for a market in risks. Risks of fluctuations in the prices of particular commodities are of this type. Merchants who have to carry stocks of, say, cotton, wheat, or copper, are subject to the risk of a change in prices. They are performing merchanting services, and will look for their remuneration to the margin between the prices at which they buy and sell, and they do not, in general, wish to run the risks of price fluctuations. There will be other people handling the commodity, manufacturers who use it perhaps, who will enter into contracts to supply their product, and who in turn do not wish to bear risks of fluctuations in the price of their raw material. The first set of people will be willing to enter into contracts to sell a quantity of commodity at some future date, while the second set will be willing to enter into contracts to buy at some future date, at a prearranged price.

For the convenience of such people the exchange handling the commodity often institutes a " futures contract," which is a contract to buy a fixed quantity of a prescribed grade of the commodity, at an agreed price, at such and such a time in the future. The contracts may be made invariably for the same period ahead, as in the case of non-ferrous metals, or contracts

may be made for any number of months, with an upper limit, as was formerly the case with cotton and wheat in this country. Such a practice is possible only when the commodity is so standardised that any sample of it can be allocated to one of a few carefully specified grades. Such contracts can be resold and may change hands many times.

In pre-war days, when the Liverpool Cotton Market was open,[1] a cotton merchant collected his stock of cotton during the autumn and early winter and " hedged " the lot, as he would term it, by making contracts to sell on various dates up to a year ahead. As he sold cotton to the spinners, he " undid his hedge " by buying contracts to receive cotton for the same dates as those on which he had contracted to sell. He bought one contract for every hundred bales of cotton he had sold from his stock, and when the time came for him to fulfil his forward contracts to sell, he did so by handing over contracts entitling the holder to the agreed amount.

There was thus a considerable trade in these futures contracts, although in the case of cotton no spinner ever thought of buying his cotton in this way, the reason being that spinners require large quantities of particular grades of cotton. If he used the market in futures contracts, he could not be certain of getting a single 100-bale lot of the grade he required. He was therefore willing to pay the merchant a small margin of profit for collecting large quantities of the grade he needed and storing it until he wanted it. The merchant performed this service at a cost of a small fraction of a penny per pound, and used the futures market for hedging, so that a change in price should not wipe out the small margin of profit on which he ran his business.

Now there was no assurance that the demand to make contracts to sell, in order to hedge merchants' stocks, would be equal to the demand to make contracts to buy. When the new crop was coming on to the market and merchants were collecting their supplies for the year, the spinners did not know the extent of their requirements, and there was likely to be an excess of contracts to sell. In such a situation professional speculators stepped in and accepted the merchants' offers,

[1] Since the above was written the Liverpool Cotton Market has been reopened. Modifications have been made to meet changed circumstances, but the principles on which it operated are substantially unchanged.

trusting they would be able to sell their contracts at a profit at a later date. Similarly, they were willing to make contracts to sell cotton they did not possess, in the hope of being able to buy later at a price which would show a profit. Such men concentrated their whole efforts in estimating the future trend of prices. If they were skilful they made a profit, and if they were unskilful they went bankrupt. On them fell the major part of the business of bearing the risks of price changes, and their profits were their remuneration for performing this service. It paid them for the expenses of collecting the necessary information and for the labour in estimating risks. Part of their profit came out of the losses of amateur speculators, who as a body always lose.

For technical reasons, which we cannot go into here, some risk had still to be borne by the merchants and spinners, but the major part was borne by the professional speculator, who was always willing to take over, at a price, a risk of the type in which he was a specialist. The merchant or manufacturer was thus enabled to contract out of most of the risk and to replace a contingent loss by a cost item.

Risks of Long-term Investment

When a decision is made to acquire some durable income-producing asset, a certain return, which can be expressed as a percentage rate of return over cost, is regarded as probable. But there are chances that the actual result, in a particular case, may differ substantially from that which is regarded as probable in the light of experience of this type of investment. The circumstances on which these probabilities depend may lie on the side of demand for the products which the assets are designed to produce. If investment is being made in equipment to produce a new product, forecasts of the demand for it may either exceed or fall short of the result achieved. If it is to make a product with an established market, there is the possibility that the demand for the product may change. Then with new products, or with new processes for making old ones, there is the possibility that it will be impossible to translate the process from laboratory scale to works production scale. The process itself may be subject to imperfectly controlled variations, a common feature where agricultural processes are

part of the production scheme. In mining, the deposits to be exploited may prove to be more or less extensive, or more or less accessible, than was originally estimated.

We can arrange investment opportunities in order of riskiness. Thus, some opportunities may offer a high chance of a moderate rate of return, slightly smaller chances of somewhat bigger and somewhat smaller returns, and little or no chance of either a total loss or a very large return. Other investments may offer a moderate chance of a very large return, little or no chance of a small return, and a high chance of a total loss. This is an investment of the hit-or-miss type so often found among mining ventures. They tend to be either great successes or complete failures, mediocre successes not being numerous.

In general, more people will be willing to take the safer opportunities for investment and, as larger quantities of the kinds of producers' goods needed for these investments are made, their marginal products will diminish and the rate of return over cost will fall. So the probable return will tend to be smaller in safer investments than in risky ones.

The investment in a collection of producers' goods to make a certain product, or an investment opportunity as we have termed it, will not offer one definite rate of return. Instead, it will offer a range of rates of return, each of which has attached to it a certain probability, which can be expressed as a vulgar fraction. The total of all these fractional measurements of probability is unity, for it is certain that one of all the possible chances must turn up. The rates of return vary, from complete loss of the whole investment, through partial return of the cost of the investment, to whatever positive rate we may think to be the maximum possible. Of all these rates one will have a bigger chance than all the others of turning up, and this we call the probable rate.

Risks of Making Loans

To make these investments, businessmen and firms will require loans and, without anticipating our discussion of the price of loans or interest, we can classify loans according to risk involved in a manner similar to our classification of investments. Even where loans are made at an agreed rate of

interest, the chances of interest being paid as it falls due and of repayment being made on the due date are not everywhere the same. Furthermore, loans are made on terms which leave the annual interest payment conditional on the success of the venture.

A glance at a few pages of the *Stock Exchange Year Book* will establish two facts about the rates of return yielded by securities. First, the same company will be obtaining some loans on very different terms from those on which it obtains others. On some of its capital it may be paying little more than the rate obtainable by purchasing Government securities, for some it may be paying 2 per cent. above the " gilt-edged " rate, and on the remainder, if it is a successful company, it may be paying a very much higher rate of return. The proportion of capital falling into each of the various classes will vary from one type of enterprise to another.

Secondly, we shall find large differences in the dividend rates paid in different types of enterprise. A careful examination of the records of a number of companies of each type over a series of years may show that in those types of enterprise, where relatively large dividends are paid by some companies, the number of failures or of companies which manage to keep alive, but which rarely pay anything to shareholders, will also be higher than normal. If, in a particular line of production, a large proportion of cases of unusually high rate of return is offset by a large proportion of cases of small or zero return, it may be that the average return is no higher than that obtainable elsewhere.

The explanation of both these types of difference is to be found in the differences in risks borne. The shares issued by a public company fall into three general classes, though each class may have several subdivisions. First, there are the debentures which bear a fixed rate of interest. The payment to the debenture-holders is a contractual payment, and the claims of this class of shareholder have preference over those of the rest in the event of liquidation. They can force the company into liquidation if their interest is not paid, but, on the other hand, they have no share in any increase in the prosperity of the company.

The next type of share is the preferred ordinary. Again, it

generally bears a fixed rate of interest, though cases of rights to a limited participation in the balance of profits remaining after all contractual payments have been met are not unknown. Sometimes dividend rights are cumulative, so that if distributable profits in one year are insufficient to pay the preference dividend, the amount unpaid is carried over as a charge against future profits. Nominally, at least, the claims of preference shareholders have priority over the third class in case of liquidation, but this right is frequently worthless, for by the time the position of a company has become so hopeless that winding-up is contemplated, there is nothing left for the preference shareholder. Indeed, experience has shown that the position of holders of this type of security is weak. They do not share in the large profits of good times, and if the venture is unsuccessful they have often suffered as severely as those who would have benefited most if the venture had succeeded.

For the third group, the ordinary shares, there is no guaranteed rate of return, and the holders may be described as the residuary legatees of the productive process. Their claims are not considered until all wages, rents, payments for materials, stores and replacements and all fixed interest payments have been met. If the concern succeeds, their reward may be high, if it merely keeps alive they will get little or nothing. The rate of dividend paid on ordinary shares will consequently vary with changes in general business activity, although the policy of paying out less than the circumstances permit in good years, in order to create reserves to augment dividends in bad years, modifies the effects of the trade cycle to some extent. But besides receiving the balance of net receipts apart from the sums put to reserve, the ordinary shareholder bears the residue of the risk, which has not been eliminated by organisation, or turned into a cost by insurance, or by transference at a price to professional risktakers. The question which the prospective lender should decide is whether the prospects of return are an adequate remuneration for the risks which he has to bear.

The demand for loans will therefore be stratified in a fashion which is closely related to the risks of the investments in producers' goods for whose purchase the loans are needed.

Supply of Loans not Homogeneous

The supply of loans will be stratified also. The supplies affected will primarily be those available for long-term borrowing, though there will also be specialised pools available for use in the speculative markets in both commodities and securities. The funds available for purchase of long-dated and perpetual securities consist in part of funds whose owners seek for security first and income second. At the other extreme there are those who are willing to undertake a large number of very risky loans in the belief that, by careful selection, a sufficient number of successful ones can be picked to give a greater return on all the loans taken together than if a selection of safer projects had been made. In between there are people of varying degrees of caution. The demand for loans, or what is the same thing the offer of securities, is also graduated to meet different tastes in riskbearing, and rates of return will be established which will give equilibrium in each market separately.

The funds available for a particular type of loan do not remain a constant portion of the total of loanable funds. At times buyers of securities as a whole are more optimistic than at others, and the prices of more risky securities are bid up so that the rate of return on the price paid falls. At times when they are more timorous, funds will turn towards safer securities, and although in such a period even the prices of gilt-edged securities are likely to be falling, they will not fall so drastically as the prices of the more speculative ordinary shares.

We must not therefore apply, too literally, the rule that the return to similar units of a factor of production should be the same in the long run. The return to the purchase of securities of all types would be the same if funds could flow freely from one market to another. This is not the case. There are legal barriers in the rules for the disposal of trust funds, there are the institutional barriers in the necessity for banks, insurance companies and other financial institutions to have large holdings of gilt-edged securities, and, above all, there are the subjective factors controlling supply which we have discussed.

Equality of payment between units of factor only applies if movement between occupations is without barriers, and this is not the case for units of capital. So it may come about that,

over a period of time, the return on risky loans may be more or less than that on safer investments. So far as American experience is concerned, there have been several computations to show that, over periods of several years, the return to an investment in a representative sample of "stocks" (ordinary shares) is less than on an equal investment in "bonds" (fixed interest securities), the computation taking into account both income and capital appreciation or depreciation. This tendency is attributed to the over-optimism of the more speculative investor, who is so confident of his ability to pick out a larger selection of successful gambles than the average of successes, that he bids up the prices of speculative securities to excessive levels.

The tendency to equality of return on loans of different types must not, on the other hand, be under-estimated. The tendency will be to equality of average return for loans of comparable types. So if risks of default differ, then in those cases where the chances of default are high, people will wish to see a prospect of a high return if the venture is successful. Suppose that in ninety-nine out of a hundred factories set up to make glass marbles, the venture pays a return of 5 per cent. to those who have made loans to it in the shape of subscriptions for ordinary shares. The average rate of return in this case will be 4·95 per cent. as compared with an expected return of 5 per cent.

If, on the other hand, in three out of four copper-mines lenders receive nothing on their loans, then prospective lenders must see a possibility of 20 per cent. before they are willing to lend to copper-mines rather than to glass-marble factories. If the rate paid by glass-marble factories is 4 per cent., then copper-mines will be considered if they can offer 16 per cent., and the fall from 5 to 4 per cent. will presumably increase investment in copper-mines as ventures which can offer 16 per cent. but not 20 per cent. to lenders will now be able to obtain funds.

The more accurately lenders can forecast risks of different types of enterprise, the more closely will average rates of return on loans for all enterprises tend to equality.

Dividend Rates and Yields

So far in this chapter we have talked in terms of making loans to firms which wish to increase investments, and have

said nothing of the market in securities or titles to old loans. It is not inevitably the case that, if at a given time a buyer of brewery shares would generally get 5 per cent. on his purchase, it would be possible to float a new brewery company and issue shares on the basis of a prospective dividend of 5 per cent. The risks which a new venture must face and those which concern a prospective purchase of shares in an established company in the same line of production are not the same. The fact that a company has reached the dividend paying stage shows that it has avoided certain dangers which the new venture has still to meet. The yields on the prices of securities of old-established companies will therefore differ much less than the prospective dividend rates it is necessary to offer to get people to subscribe to new issues. The yield, or the dividend calculated as a percentage of the market price of a share instead of on its face value, does not depend so much on the size of the dividend as on the chances of that dividend being maintained in the future. If one company is paying £5 on one hundred pound shares and another company is paying £20 on the same amount, and *the chances of the two companies continuing to pay these dividends are considered the same*, then the yields on the purchase of the two lots will be the same, the price of the second being four times the price of the first. The chance of any firm, irrespective of the industry to which it belongs, being able to maintain a particular rate of dividend is always largely dependent on changes in the general level of business activity. Some industries, it is true, are less affected than others by either boom or slump, yet to a considerable extent the fortunes of all wax and wane together. Changes in the rate of interest affect them all, and they are each affected by changes in the values of the marginal products of the capital goods which the investment represents.

The reality of this distinction between the risks attaching to the initiation of an enterprise and those attaching to its conduct is illustrated by the case of the Suez Canal Company. No comparable adventure had been undertaken in modern times and, though there was little doubt of the profitability of the canal once it was cut, there were serious doubts as to whether it would be possible to cut it. In addition, the Egyptian Government gave a concession for a term of years at the end of

which the canal was to become the property of Egypt. The Canal Company had therefore to make arrangements for the repayment of capital, and did this by annual drawings of shares for redemption. If no other provision had been made, a shareholder, who subscribed for shares on issue and had them drawn for redemption after a few years, would, for a relatively small payment, have borne the risk of loss of his loan through failure to construct the canal. So it was arranged that a fixed dividend of 5 per cent. should be paid on all shares and, in addition, a variable dividend depending on the profitability of the venture. When a share was drawn, its right to the 5 per cent. ceased, and it was replaced by a non-redeemable share with a right to equal participation in the variable dividend.

The Function of the Theory of Profit

In this examination of the risks of investment, an attempt has been made to break with the traditional treatment of profit, where part of the reward to capital was labelled interest and was supposed to be paid at a uniform rate on all loans and investments, no real distinction being made between these two terms. Any payment to capital over and above this amount of interest was termed profit, and was regarded as a reward to the function of riskbearing which was elevated by some to the status of a factor of production.

In fact, all economic activities are attended by risks, and we could apply these principles of probability theory in an examination of the price of any factor of production. It so happens that there is most scope for the use of probability theory in an examination of the problems of differences in the rates of return over cost to investments and in rates of return on loans of different types. Logically, there is no reason why profit theory should not cover all explanations of differences in prices of similar units of commodity or factor which would bring rent theory and monopoly theory within its scope, but it is much more convenient to confine it to a study of those differences in the prices of similar assets, which are attributable to the risks and uncertainties that attend economic activities. The presence of risks affects demands and supplies of factors of production of all producers' goods and of loans. Since our discussion has so far run primarily in terms of durable pro-

ducers' goods and loans, we will extend it to the pricing of non-durable producers' goods.

Spot and Forward Commodity Prices

In our studies of price determination earlier in this book, we were dealing with prices at which goods were available for immediate delivery, but in the real world there may be several prices for the same commodity according to the time at which delivery must be made. There will be a price for a particular grade of a commodity deliverable at once, and this is termed the spot price. It is true that in certain markets, cotton for instance, the term " spot " has been given a wider meaning, but we will use it in the narrower sense here. Then there are forward prices for delivery one month, three months, a year later. There is also the term " future " price, but this has a specialised meaning and is related to the price of one or more particular forward contracts of the exchange, and so we shall retain the more general term " forward."

The spot price depends on demands and supplies now, but we must remember that present demand includes demand to buy now, to store and consume later. Forward prices are based on estimates of what demands and supplies of the commodity are likely to be at the date in question. The cases where transactions for forward delivery commonly take place are those of commodities which can be stored, and so there is a relation between spot and forward price. If we can find a commodity which cannot be stored over the time period involved, then there is no necessary connection between spot and forward prices.

If an increase in demand is expected in the future, the forward price will rise and the existing commodity will be stored instead of being consumed now. The spot price will rise also, even if present demand is unchanged. Also we must remember that if a commodity is to be stored, prices must be such as to induce people to store it ; that is, the forward price must exceed the spot by an amount at least equal to the cost of storage for the period involved.

Now let us work out the relationship between the two prices on a series of assumptions. Suppose we are dealing with a commodity like an annual crop, where the harvest is due in six

months' time. Conditions of demand, we will suppose, are expected to remain the same for the next six months, and there is enough stock to last till harvest at the current rate of consumption. The next harvest is expected to be the same as the last. Commodity must be stored for periods up to six months if consumption is to remain steady, and so forward price must be above spot price by an amount sufficient to cover storage costs. A seller would clearly be willing to accept spot price plus four months' storage cost for a contract to deliver four months hence, but would a buyer be willing to pay that price if it corresponded to his estimate of the price four months hence? He would not. His estimate of the future price may be too high, and so his offer will equal his estimate of that price less a margin for risk, and if this reduced price is not less than spot price plus four months' storage, a bargain can take place. If the estimated price four months hence, minus the premium for risk, gives a figure which is greater than spot price plus four months' storage charge, or, to put it another way, if the estimated price exceeds spot price plus storage charges by an amount greater than the risk involved is considered to justify, then the spot price will be bid up as buyers acquire the commodity, now, for storage. Equilibrium will have been reached when the forward price is less than the expected price four months hence by a margin which reflects the estimate of risks involved and the forward price exceeds the spot price by the amount of carrying charges. So, for periods up to six months, there will be a rising scale of prices, the rise corresponding to the growing costs of storage, and each price being slightly less than the price which is expected to rule at that time. These prices will show a " contango," the term for the price relationship when the forward price exceeds the spot.

If conditions are such that the price of the new crop in six months' time is expected to be the same as the price today, the six months' forward price will be slightly less than the spot price, the difference being a margin to cover risks of the estimate being wrong. Spot price will exceed the forward and there will be a " backwardation."

Now suppose that a catastrophe destroys some of the current stock, but leaves the prospects of the next harvest untouched. The six months' price will be unaffected, but stocks are not big

enough to last six months at current rates of consumption. Spot and forward prices for periods less than six months will rise, until the rate of consumption has been cut down to a level which can be maintained for six months and the relation—spot plus storage charge—will hold between the spot price and each forward price. As spot price has risen, the backwardation on six months will have increased.

A prospect of a short crop will result in a rise in the six months' price, which may show a contango, and the spot price will rise to cut down current consumption and make present stocks last more than six months. If people expect a larger harvest, or the prospects for the future look less promising, their estimate of the price for the next crop will fall, and they will probably use a larger margin for risk, so that the six months' price will show a larger backwardation. Spot price will be unaffected if current stocks are considered large enough for six months' consumption only, but if there are stocks surplus to this requirement, spot price will fall also. If for any reason it is expected that stocks will have to be held over a period of time, the spot and forward prices for that period will show a contango, as spot will fall sufficiently below the expected price at the end of the period to create a contango equal to the cost of storage.

In this way we see how considerations of risk enter into the structure of commodity prices, thus affecting the demand for commodities, and since commodities in which there is a forward price are usually raw materials, the volume of employment offered is also affected.

Risk Estimates and Demands for Commodities and Factors

Present demands for factors of production and commodities arise largely because people are catering for other demands which they expect will occur in the future. Land is being ploughed today because people will have a demand for bread or cotton shirts two years hence. The expenditure undertaken now depends on estimates of the size of that demand.

But it is not merely the estimate of the size of the demand which matters; estimates of the reliability of estimates of demand are also important. The more confident people feel about the reliability of an estimate, the greater the provision

they will make in the light of it. We have seen that in the capital market the amount of capital which will be provided in response to a certain probable yield depends, not only on the size of the prospective return, but also on the degree of risk that the yield will not materialise. More capital will be subscribed where there is a nine-tenths chance of getting the probable return of 5 per cent. than if the chance of getting that return is one-half.

If the present price of a raw material is £20 per ton, present stocks are expected to be used up during the coming year, and the expected price a year hence is £19; then, if there is great confidence in the reliability of this estimate, the twelve months' forward price may be £18, but if the future is obscure and the basis of estimate more doubtful, the backwardation will be larger and the forward price may be only £15. At this figure less provision for the future will be made than if the forward price was £18, because it is the practice of people making plans for future production of materials to sell the whole or part of their expected output forward some time before it is actually in existence. Output is therefore reduced until the anticipated marginal cost is equal to the forward price. Again, considerations of risk influence output.

Risk and Demand for Factors of Production

It is clear, from the above discussion, that the effective demand for producers' goods, whether the permanent equipment of production or the raw materials and semi-finished products necessary to keep the fixed equipment working, depends, not only on the levels of prices expected to rule in the future, but also on the degree of reliance which can be placed on those estimates. It may be that if existing conditions in the copper market remain as they are today, it is reasonable to suppose that price in three months' time will equal today's price, but there is the risk that demand may change, that the political situation in a large producing country may lead to a fall in supply, or that in a large consuming country to a fall in demand. So it may be prudent to take a less optimistic view and to make plans on the assumption that the price three months hence will be £20 a ton less than it is today. The more uncertain is the future, the fewer the plans which businessmen will be willing to make for the future.

This in business jargon is what is termed the state of confidence. If businessmen are confident in the stability of the underlying conditions of demand and supply as they exist and they consider that any change will be in their favour, they will enter into commitments for the future, and this means that either they or those to whom they give orders will increase their demand for labour *now*. Increased expenditure on producers' goods increases the incomes of the factors of production engaged in making such goods. Production income is greater than it would be without this expenditure, and provided it is not a question of replacement of depreciation there will be an increase in earnings also. We have, in these changes in confidence, i.e. changes in the degree of risk attached to estimates of the future, the most potent factor inducing changes in the volumes of income and employment.

Here, however, we are stepping outside the Theory of Risk into the field of the Theory of Expectations, which is concerned with changes in estimates of future risks.

Further Reading

Knight, F. H.: *Risk, Uncertainty and Profit*, Parts 1 and 3.
O'Brien, G.: *Notes on the Theory of Profit.*

More Advanced Reading

Beddy, J. P.: *Profits.*
Keynes, J. M.: *Treatise on Money*, Chap. 29.

Chapter XVIII

THE SUPPLY OF MONEY

SINCE we suspect a relation to exist between the supply of money and the rate of interest, it is essential that we should investigate the conditions which underlie the supply of money. The generalisations we shall make will not have the universal validity of some of those at which we arrived in our discussions of diminishing returns or marginal cost. On the whole, much more attention will have to be paid to the character of the existing institutions, so that many of our observations will apply only to Great Britain; others to any country whose banking system is built on the same general lines as ours; few will be of universal application. We can begin, however, with one generalisation which is always valid, a definition of money.

What is Money?

Looking around the world, we find an amazing variety of objects which are termed money. In the more sophisticated countries pieces of paper printed with intricate designs, difficult to copy, are popular. But this is a fashion which has spread a good deal in the last century. Before that, coins, which bore stamps testifying to the weight and quality of the metal in them, were in common use. Then as now, other pieces of metal of various sizes and bearing various designs were used, but these differ from the first kind, in that they will buy more unstamped silver, nickel or bronze than they themselves contain. The first kind, which are always gold nowadays, exchange for their own weight of uncoined gold. In less sophisticated countries cattle, pigs, bars of salt, sticks of tobacco, cowrie shells, or even boulders of an uncommon sort of stone may be generally acceptable in payment of debts. This varied assortment of objects has one characteristic in common. *They are commonly accepted*

in the community which uses them in payment for goods and in settlement of debts. This is a function any commodity which is to be used as money must fulfil. It must serve as a *medium of exchange.*

In less sophisticated communities the commodity may be one which is highly valued for some intrinsic property: its decorative value, its magical properties, or even its popularity in consumption—these may contribute to its general acceptability. There is always the comforting assurance that if there is no opportunity to use it as money, it will look very nice strung round the neck or will be good to eat or to smoke. In more sophisticated communities these qualities may not be regarded as advantages, and suitability as a medium of exchange is the sole criterion for acceptability. Such is the case with our notes and token coinage. The more important the place of exchange in the economic life of the community, the more does the medium of exchange function overshadow any intrinsic usefulness of the commodity.

In addition to acting as a medium of exchange, money generally provides a unit in terms of which the prices of all other goods and services are measured. It is then said to serve as *unit of account* as well. The two functions are usually performed by the same thing, but this is not invariably the case. In certain African communities prices are measured in terms of cattle, although cattle rarely change owners. In our own country certain prices are conventionally measured in guineas, although the guinea coin has long ceased to exist.

A further stage in increasing sophistication is that people wish to store wealth in a generalised form, as distinct from stores of useful objects. This can be done by storing the commodity which acts as the medium of exchange, provided of course that it is durable, for a commodity which is difficult or expensive to store safely will not be attractive as a store of value. Furthermore, once the idea of using a commodity as a store of value has been adopted, there arises the question of the constancy of the value of the money commodity in terms of other commodities. People will not wish to see their store suddenly lose most of its purchasing power over other commodities because the commodity has become more plentiful. So a commodity whose supply is more inelastic than that of any

other has commonly been chosen. Ideally, then, the commodity should be sufficiently plentiful to enable the exchanges people want to make to be effected, but the supply should be more difficult to increase than that of any other commodity. In addition, it should be storable and easily divisible, so that payments can be made in amounts, however small.

Kinds of Money

It is convenient, in a community which practises a complicated system of exchange, to have different kinds of money. There will be a unit like the pound sterling, the franc or the dollar, and multiples and submultiples of it. The unit may be a coin containing metal of intrinsic value equal to its face value, like the pre-1914 English sovereign, or it may be a paper note, which could, on certain terms, be exchanged for a stated quantity of gold, like the British pound note between 1925 and 1931. Such coins as the bronze penny and the silver-alloy shilling were acceptable, even in the days of the sovereign, because it is convenient to have coins of small denomination for the purpose of making small payments. Such coins are called "token money," because their intrinsic value is less than their nominal value, in contrast to " full-bodied money " like the pre-1914 sovereign. The Bank of England pound note is also token money, for its intrinsic value is negligible, but because of its usefulness people are willing to accept it.

These types of money are issued under governmental authority, and though governments cannot secure a monopoly of the issue of money, they can endow the money they issue with the special quality of being "legal tender." A creditor is obliged by the law of this country to accept pound notes in payment of a debt, but in the case of the silver and copper coinage, there are limits to the amounts which he is obliged to accept, and such money is hence described as limited legal tender. The fact that they have the authority of the Government behind them adds greatly to the acceptability of such forms of money. People will have to accept such money as wages even if they do not like it, and a good deal will be used, although there is nothing to prevent the owners of goods from refusing to sell them in exchange for legal tender money, if they think they can get some other kind they prefer.

A certain amount of inconvenience attaches, however, to any of these forms of common money, as we may call them. If large payments have to be made, there is the trouble of counting and watching that no worthless imitations are included in the sums received. When payments have to be made between persons in different parts of the country, there are the costs of having money transported and the risks of loss in transit. A certain amount of trouble attaches to the safeguarding of the sums, which even persons of very moderate means possess, if they use only common money. A demand therefore arises for the services of institutions which will take care of stocks of common money and make payments from them as and when and where the owner desires. It is necessary that such institutions should afford a high degree of security for the funds entrusted to them and that they should be able to effect transfers at low cost. These services are performed by the banking system, and in order to perform them, it creates a further kind of money, " Bank Money."

Supply of Bank Money

The services of providing safe custody for gold and of transmitting funds for the settlement of debts between people in different countries, or in different parts of the same country, were originally in separate hands. Because of the nature of their trade, goldsmiths had to provide for the protection of their stocks of precious metals and, at little expense to themselves, they could undertake to look after the valuables of other people. On the other hand, people who were engaged in international trade would have debts falling due to them in certain towns and would have debts to pay in other towns. If they could find someone who had a debt to collect in the centre where they owed money and who was willing to sell the debt to them, then, by offsetting the two transactions, the amount of money to be carried between the two centres could be reduced. Merchants engaged in international trade thus found it advantageous to become middlemen in international debts as well, and some of them ended by giving up trading in goods and became international bankers. The aspect of banking with which we are concerned at the moment derived, however, from the goldsmith-ancestor of the modern banker.

The goldsmiths found that people who deposited gold with them did not all require it at once, and that it was possible to lend some of it to credit-worthy borrowers. So long as they kept a reserve sufficient to meet any possible calls on them from owners who wished to have their gold back again, they were quite safe in lending out the rest, and charging interest for the loan. A goldsmith who pursued a prudent policy found the practice profitable, and indeed found it paid him to offer a rate of interest to all who would deposit gold with him, instead of charging a fee for safe custody as formerly.

A further step was taken when, instead of handing gold to borrowers, the goldsmith handed a promise to pay gold. He found that he could issue more promises to pay gold than the quantity of gold in his possession could satisfy, because all the people who held his promissory notes did not ask for them to be paid in gold. As the goldsmith's reputation grew, his notes passed from hand to hand in settlement of debts. The skill of the goldsmith, or the banker, as he has now become, lay in his judgment of the quantity of the notes it was safe for him to issue. The more he issued the greater his profits from interest, but the greater the risk that they would get into the hands of people who did not care for them and would demand gold in exchange. Also there was always the possibility that his customers would lose their confidence in him and come clamouring for their gold, which he would be quite incapable of finding.

When the transition from goldsmith to banker was complete, the bankers, instead of issuing a single note for whatever sum had been deposited or requested as a loan, printed notes of fixed denominations, and the banknote proper evolved. We have, then, two kinds of money in use—gold and notes, and one is convertible into the other. There will be a steady stream of business, some people cashing notes in gold, others giving gold for notes. What the banker had to fear was a change in the acceptability of his notes, so that demands to repay notes in gold would exceed deposits of gold in exchange for notes. He had therefore to carry a stock of gold big enough to cover any loss of popularity of his notes, which his experience taught him might occur.

Bank Deposits

So it is with the contemporary descendant of the goldsmith, the deposit bankers. Here again we have two different kinds of money, common money consisting of notes and coin, and bank money, and one is convertible or exchangeable into the other. Bank money is a less tangible form of money than any we have yet encountered. It consists of entries in the ledgers of the banks. The banks can create as much bank money as they choose by making entries in their ledgers, so that at first sight this kind of money does not fulfil the criterion of inelasticity of supply, but we must remember that the bank is liable to exchange bank money for common money on request, so that its desire to issue bank money, in order to make profit from the interest it charges for doing so, is restrained by the necessity for having sufficient common money to satisfy those who demand it in exchange for bank money in their possession.

The supply of bank money comes into existence in two ways. People deposit common money with the banks and receive an equal amount of bank money in exchange. Their reasons for doing this are that a stock of common money must be preserved against loss and theft, but an entry in a bank ledger requires no such care; also it is more convenient to make many payments by transferring bank money than by handing over notes, for no questions of miscounting or of forged notes can arise, loss in transit causes merely minor inconvenience, and a record of the transaction is automatically created.

The second way in which bank money is issued is by the granting of a loan by the bank to a customer, and in this country the usual method is for the bank to grant an overdraft. The customer who has received an overdraft is in the same position as if he had a deposit of the agreed sum standing to his credit, but in fact the banks do not count overdrafts granted but unused in their total of "deposits." The true total of bank money available at a given time is therefore the total of "deposits" and unused overdrafts.

The most usual way of transferring bank money from one person to another is by means of a cheque, which is a formal instruction to a bank to pay to a specified person or to "Bearer." There are other means, such as Bankers' Orders and Bills of

Exchange, but for the present we shall talk as if the cheque was the sole method. Each day a bank receives from its customers quantities of notes, coin and cheques, which have been drawn in favour of the customer by a third party who owes him a debt. The notes and coin are entered to the credit of the customer who pays them in, and the total deposits in the books of the bank are increased by that amount. The value of the cheques is also entered to the credit of the customer, but cheques transfer bank money, and although they increase the amount standing to the credit of the customer who pays them in, they do not necessarily increase the total deposits on the books of the bank.

A Single-bank System

To simplify the problem, we will first suppose there is only one bank in the country. Then a cheque drawn in favour of John Jones of Newcastle by Harry Smith of Coventry will, when paid into the Newcastle branch, increase the deposit to the credit of Jones and decrease the deposit at Coventry to the credit of Smith by the same amount. The total deposits of the bank will be unchanged ; only their ownership has changed.

There is, however, a circumstance in which total deposits can have changed. If Smith had no deposit to his credit, but had arrangements with the Coventry branch for an overdraft, there would be no decrease in deposits at Coventry to offset the increase at Newcastle, and the total deposits of the bank would have increased.

But all of us do not have bank accounts, and those of us who have require common money to make certain payments, so there will be a daily flow of notes and coin out of the bank. The flow in and the flow out will not cancel out every day. If it is the custom to pay wages on Friday, the outflow on that day will be larger than normal, but probably partly compensated by a big inflow on Saturday, when people spend a large part of their wages. If there are customary days for paying rents, people who have no bank accounts hoard common money until rent day, when it will return to the banks. A certain reserve of common money is therefore necessary to allow for these fluctuations, because if a bank is to retain the confidence of its customers, it must always be in a position to exchange bank money for common money on demand.

Quantity of Loans and Demand for Common Money

We can now see where the check arises on the ability of the bank to create unlimited quantities of bank money by making loans. If a bank makes loans to customers, their demands for goods and services are increased. The bulk of the loans are made for business purposes, and are spent by the borrowers on paying wages and buying materials, so that directly or at one remove the loans are likely to result in increased payments to people who do not own bank accounts and who will therefore demand common money. Increased loans will therefore lead to an increased demand for common money and, if the bank is too free with its loans, it will find that its stocks of common money are not sufficient to absorb the fluctuations in the inflow and outflow of notes and coin.

The Cash Deposits Ratio

The bank is therefore likely to set some limit on the amount of loans it makes. Now, the real problem is the change in relative demands for bank money and common money, but the banks are most concerned about their deposits, which are liabilities to pay common money to their customers. They can always withdraw unused overdraft facilities. It has therefore been the practice of banking systems to set either legal or conventional standards for the proportion between their holdings of common money and deposit liabilities to customers. In the British banking system the conventional proportion until recently was that holdings of common money should represent 11 or 12 per cent. of deposits. In 1946 the proportion was reduced to 8 or 9 per cent. by agreement between the banks.

At this point we must allow for the complication which arises because there is not one but a number of banks. Each individual bank when it makes loans has to face the probability that the bank money it creates will in part get into the hands of the customers of other banks. The recipients will pay the cheques into their own bank, which will ask the issuing bank for common money in exchange for its bank money. So if in the example we used earlier Smith is a customer of Lloyds, who grant him an overdraft, and Jones is a customer of Barclays,

then before the transaction is finally settled Lloyds will have to pay Barclays the appropriate sum in common money.

There are of course many such transactions in either direction between each pair of banks each day, and every effort is made to offset transactions. Thus in each town the cheques drawn on the branch of one bank in the town and paid into the branch of another bank in the same town are offset against each other. Then there are the large towns which act as clearing centres for districts, and all cheques paid into banks within that district are offset. Finally, there is the London Clearing House, with its three divisions of town, metropolitan and country, where the net balances owing by one bank to another are determined. Any balances which must be met to settle the day's transactions are discharged by cheques drawn on the account each of the banks has at the Bank of England. The balances in these accounts are treated as if they were notes, as the bank can always get notes from the Bank of England on demand.

It is obvious that if one bank grants loans on a larger scale than the others, its stock of common money is going to drain away into the other banks, even if there is no increased demand for it from the members of the public. The agreed ratio of cash to deposits therefore serves the further purpose of keeping the banks in step. Because they all observe this ratio they create loans at the same rate, and the chances that the net flow of cash from one bank to another will be zero over short periods of time are increased.

Relative Quantities of Bank and Common Money

The relation between the quantities of common money and bank money in a community will depend primarily on the habits of its members. If the community has little faith in the banking system, or if for some other reason, such as " black market activities " or desire to evade income-tax, a larger part of the transactions of the community are settled by cash payments, the demand for cash will be larger and the demand for bank money smaller in relation to a communal income of a given size than if these factors were different.

Different income groups in a given community have different habits in the matter of the proportion of their resources which they carry in the form of notes, so that changes in the distribu-

tion of incomes between social groups can alter the demand for one or the other form of money, even if the communal income is unchanged.

Given the demands of the community, the demand for common money on the part of the banks, in relation to the quantity of bank money it has issued, will depend on its structure and efficiency. So, given the habits of the community and the form of the banking system, the amount of bank money created by the banking system will depend on the amount of common money in existence, and we must enquire how that is controlled.

The Supply of Common Money

So essential to the efficiency of the production system is the existence of a trustworthy currency that control of its issue is one of the first functions undertaken by the State. The task is commonly delegated to a Central Bank, which does little or no banking business with the general public, but acts as banker to the Government and to the other banks. As a result of these activities, it also controls the supply of money. The powers of the Bank of England, the Central Bank of the United Kingdom and Northern Ireland, have long been strictly regulated by Parliament. Its Central Bank functions developed spontaneously through its monopoly of joint stock banking in the City of London, and its position as the Government banker. The Bank Act of 1833 gave legal recognition of this fact and regulated its performance of those functions. Since then Parliament has fixed a limit to the number of notes the Bank can issue for the purpose of lending to the Government or for the purchase of Government securities. For every note issued beyond the Fiduciary Issue, as this quantity is called, the Bank must hold gold which it has valued at a fixed price, now 248*s.* per fine ounce.

There was a time when the Fiduciary Issue represented only a small part of the note circulation. In the later 1930's the Bank held some hundreds of millions of pounds in gold, and issued a corresponding total of notes over and above the Fiduciary Issue. Since the early days of the war the Fiduciary Issue has greatly increased and the proportion covered by gold has been negligible.

The Bank of England Return

The Bank of England, even in the days when it was owned by private shareholders, was not under the normal obligations of a public company to publish accounts. Instead it has had to publish a weekly return of certain assets and liabilities. The form of the Return has altered slightly from time to time, but it goes back in substantially the same shape to the Bank Charter Act of 1844.

The primary activities of the Bank are to issue currency and to act as banker to the Government and the commercial banks. The Return is in two parts, which reflect these activities: the Issue Department looks after the issue of notes and the Banking Department keeps the accounts of the Government and a number of semi-official bodies, the commercial and private banks, and a small number of private customers.

THE BANK RETURN FOR JAN. 8, 1958.

Issue Department

	£m.		£m.
Notes issued:		Government debt .	11·0
In circulation .	2,033·7	Other Government securities . .	2,085·2
In banking dept. .	66·7	Other securities .	0·8
		Fiduciary Issue .	2,097·0
		Gold coin and bullion	0·4
		Coin other than gold .	3·0
	£2,100·4		£2,100·4

Banking Department

	£m.		£m.	£m.
Capital	14·6	Government securities		207·0
Rest	3·5	Other securities:		
Public deposits .	10·5	Discounts and advances . .	19·1	
Other deposits:		Securities . .	22·0	
Bankers' . .	213·4			41·1
Other accounts .	75·2	Notes . . .		66·7
		Coin . . .		2·4
	£317·2			£317·2

The first item of the assets of the Issue Department, about £11 millions of Government debt, is a long-standing loan to the Government, most of which dates from the foundation of the Bank; but the main backing of the Fiduciary Issue is the ordinary Government securities which the Bank buys when the issue is increased and sells when the issue is diminished. The holding of gold is trivial compared with pre-war days, when the Fiduciary Issue was £200 million. Against these assets the Bank issues notes which are mainly in the hands of the public or of the commercial banks, and the balance forms a stock in the Banking Department, which must have a reserve to meet the demands of its customers for notes.

The second part of the Return deals with the Banking Department. Here the liabilities begin with the capital subscribed by the shareholders, or the proprietors as they used to be called. The next item, "the Rest," never falls below £3 million; anything above this amount is regarded as distributable profit. Deposit liabilities are now separated into three items: the first, public deposits representing sums standing to the credit of the Exchequer, Commissioners of the National Debt and the Savings Banks. The next item represents the balances of the English banks, whose principal activities are in England. By means of cheques on their accounts with the Bank of England, they settle any net balances of transactions among themselves. It is their custom to treat their balances at the Bank of England as equivalent to notes in their tills and, since they keep their credit at the Bank and the money in their tills roughly equal, the total of bankers' deposits is a useful indicator of the amount of bank credit available for business. The final item on the liabilities side represents the balances of the other clients of the Bank, Dominion and foreign banks of all kinds, British financial and banking houses other than the commercial banks, colonial governments and municipalities, numerous semi-official bodies and a few private customers.

Now let us turn to the assets side of the Banking Department account. The first item, Government securities, consists of "Gilt-edged" securities, which the Bank buys and sells on the Stock Exchange; Treasury Bills, which are promises by the Exchequer to pay a fixed sum on a certain date, usually three months after the date on which they are sold by tender; Ways

and Means Advances, which are loans to the Government to help it bridge over the gaps which occur when expenditure is continuous and revenue accrues at intervals. Securities other than Government obligations are shown separately. They may be of any type other than ordinary shares, though issues of Dominion, Colonial and foreign governments probably predominate. Commercial bills, which the Bank buys, as distinct from discounting them for customers, are included in this item. The Discounts and Advances consist of Commercial Bills and Treasury Bills which the Bank has discounted for its clients, advances made to various financial institutions and to its own customers. Finally, we have the item Notes and Coin, which is the Bank's reserve against demands for notes from the commercial banks. These are continually paying old dirty notes into their accounts and withdrawing new ones. The old ones are passed to the Issue Department for cancellation and replaced by new ones. If the banks withdraw more than they pay in, they deplete their balances with the Bank of England, and also reduce the stock of notes in the Banking Department. In the opposite circumstances both balances and the Bank's reserve grow. The ratio of reserve to deposit liabilities is termed the " Proportion".

Control of the Quantity of Money

The quantity of money is equal to the quantity of common money, plus the quantity of bank money, and the amount of the latter the banks are able to create is limited by the amount of the existing supply of common money the public is willing to leave in their hands. The existing supply of common money is controlled by the Bank of England.

It does this through the buying and selling of securities. If it wishes to decrease the quantity of money, it sells securities on the Stock Exchange, lowering the price at which it is willing to sell until buyers are found. The buyers will pay for their purchases with cheques on the commercial banks, so that a part of the Bankers' Balances is transferred to the ownership of the Bank of England.

Now two things will happen. As their credit at the Bank has been reduced while their holdings of notes have remained unchanged, the commercial banks will pay notes into their accounts at the Bank in order to restore the equality between

their balances there and their holdings of notes. The Bank of England will thus recover notes at the expense of notes in the hands of the public to the extent of half the value of the securities it has sold.

Secondly, the commercial banks, having had their cash reserves reduced by the amount of the sale of securities, will endeavour to reduce their deposits by calling in loans and to restore the 8 per cent. reserve ratio. So, as old loans fall to be repaid, they will decline to make new ones, and the volume of their loans and hence of their deposits will decline. The quantity of bank money will therefore also be reduced as a result of the sale of securities by the Bank of England.

The Bank uses this technique whenever it wishes to increase or decrease its reserve of notes, whenever it thinks the quantity of notes in the Banking Department is too small or too large in relation to the probable calls of the commercial banks upon it. The same method is used when the Government decides to vary the volume of the note circulation. If the Government decides to increase the quantity of money, Parliament authorises an increase in the Fiduciary Issue, and the Issue Department prints the necessary quantity of notes and sells them for securities to the Banking Department. The Banking Department buys securities, paying for them in cheques upon itself. The cheques find their way through the sellers' bank accounts to the bankers' deposits at the Bank. The commercial banks, finding their deposits increased, withdraw notes to equalise cash at Bank and till money, so the notes are drawn into circulation to the extent of half the purchases. The commercial banks then use their increased reserves to make more loans and the quantity of bank money is also increased.

Similarly, if the Fiduciary Issue is reduced, the process is used in reverse, and the notes find their way out of circulation into the Issue Department, where they are withdrawn and destroyed. The extra demand for notes at Christmas time is met by a temporary increase in the Fiduciary Issue and the increase and subsequent decrease in the circulation are carried out in this way.

The Quantity of Money and the Rate of Interest

There is an aspect of this technique of regulation of the quantity of money which we have so far ignored. When the

Bank of England is determined to reduce the quantity of notes in circulation, it sells securities, and to make any appreciable effect on the quantity of money the sales must be in considerable volume. The sale of a large quantity of Government bonds is likely to cause a fall in the price of such securities ; that is, a given annual income will be purchasable for a smaller capital sum than formerly. The rate of return to investment in Government securities will therefore be raised.

Now we must assume that formerly the market in investments was in equilibrium; that is, the personal estimates on the part of investors of the relative values of securities was the same as the relative market prices. In these circumstances there would be no net tendency for investors as a whole to try to sell one type of security in order to buy another. This state of affairs has been upset by the fall in the prices of Government bonds which makes it possible to earn more income by investing £100 than was formerly the case. Investors will therefore be more willing to hold Government bonds than they were before, and will sell other securities in order to buy such bonds. The prices of other securities will therefore fall also, and yields on investment will rise until a new equilibrium has been reached and investors' valuations and market quotations are in harmony once more.

The effect of decreasing the quantity of money has therefore been to raise the rates of return to purchases of securities or interest rates, as we will provisionally call them. In a similar fashion we could show that the effect of an increase in the quantity of money would be to raise the prices of securities and so lower interest rates. It would therefore appear possible to control the quantity of money and control the rate of interest by one and the same means.

Further Reading

Cole, G. D. H.: *Money, Its Present and Future.*
Crowther, G.: *An Outline of Money*, Chaps. 1 and 2.
Robertson, D. H.: *Money*, Chaps. 1, 3 and 4.
Sayers, R. S.: *Modern Banking.*

More Advanced Reading

Keynes, J. M.: *Treatise on Money*, Chaps. 32–35.

Chapter XIX

CHANGES IN THE QUANTITY OF MONEY

IN the previous chapter we discovered one effect of changes in the supply of money. As a result of the mechanism of changing the supply, interest rates rise for a fall in the quantity of money and fall for an increase in the quantity of money. It might appear that this was a result of the particular mechanism, but, as we shall see later, any mechanism for changing the supply of money will produce these effects, though the speed with which the effects of change are felt will not be the same in every case.

Quantity of Money and Commodity Prices

There is another effect of varying the supply of money on which we have not touched: its effect on the prices of commodities. Let us suppose that the Government printed more pound notes, making a present of ten to every citizen, and consider what the effect would be. Some people would hide some or all of the notes in a secret hoard, and their effect on prices would be nil; many would spend some or all of them on things which previously they considered they could not afford. Demands for certain commodities would increase, and there would be a tendency for the prices of these commodities to rise. How much they would rise would depend on the shapes of the curves of their marginal costs of production in the vicinity of the current output. We said earlier that marginal cost curves were commonly of a flat U-shape, so that if existing output should lie somewhere in the middle of the flat portion, a considerable increase in output might take place without any significant rise in price. On the other hand, if outputs were such that marginal costs were rising sharply, there might be a considerable rise in price for a negligible increase in output.

At first any price rise would be confined to the commodities which were in increased demand, but increased output of these

would lead to producers bidding up the prices of factors of production used both by them and by other industries. In this way the price rise will spread to other commodities not originally affected. Wages will tend to rise, first as a result of an increased demand for labour, and secondly when workers experience an increase in the prices of the things they buy and consequently a rise in their cost of living which provides the incentive for demands for wage increases.

The reaction to the increased supply of money would depend on the availability of unemployed resources in the economic system. If there are unemployed factors of production which can be drawn into employment, the prices of factors will not rise to the extent which will occur if an expansion of the output of any one commodity necessitates the attraction of units of factor already employed elsewhere.

Clearly the nature of the first reactions, the impact effects we can call them, of the increase in the quantity of money depends on the identity of the recipients. If, instead of paying £10 to every citizen, the Government distributed the same total sum in repayment of Excess Profits Tax, we should expect a different impact effect. The companies receiving the refund might invest the proceeds in new machinery, and the prices of such things as machine tools and electric motors would rise. On the other hand, the companies might distribute the refund as increased dividends to their shareholders, who, being on the whole of the wealthier part of the community, would not spend their windfall in the same way as the poorer majority, who would receive the bulk of the distribution if it were made on an equalitarian basis. We should expect an increased demand for motor-cars, luxury goods and continental holidays. Of course, gradually as the increased demand for motor-cars led to increased demand for labour to make them and to increased wage payments, the flood of new expenditure would seep into all channels, but its effect is likely to be greatest where its impact is first felt.

For our final hypothesis let us return to the first case, and suppose that every recipient of £10 paid it into a savings account at a bank. The commercial banks, finding their cash reserves increased, would be willing to increase their loans, and to do this it is probable that they would have to lower the rate

of interest charged for overdrafts. With a reduced cost of borrowing, it will now be profitable to hold investments which formerly it was not worth-while holding on borrowed funds. Some of these investments will be securities, and an increased demand for securities will lead to a rise in their prices ; that is, to a lowering in the rates of return they afford. On the other hand, businessmen might borrow money to expand production and the employment of labour.

There is not the same degree of certainty of prediction in this case as in the first and second cases we examined. The acquisition of increased cash reserves puts the banks in a position where they are able to increase credit. It does not follow inevitably that borrowers will find the terms sufficiently attractive to induce them to increase their borrowings, and if they do not borrow more, the effect will be just the same as if the recipients of the extra notes had hoarded them.

The Quantity Theory of Money

Let us now see how what we have learned about the relationship between changes in the quantity of money and changes in prices accords with the classical explanation, the so-called Quantity Theory,[1] which expressed the relationship as $M.V = P.T$, where:

M the quantity of money;
V the number of times a unit of money changes hands per annum on an average;
T the number of transactions per annum;
P the average price per transaction.

This equation is an identity stating the indisputable fact that the total volume of cash transactions, $P.T$, is equal to the total of cash payments. If all payments were made by means of cheques drawn on bank accounts, then $M.V$ would be equal to the total of cheque clearings, or since in British banking statistics this term is restricted to cheques cleared through the London Clearing House and the eleven provincial clearing

[1] The Quantity Theory of money in its original crude form, which stated that the level of prices was dependent on the quantity of money, is attributed to Bodin, *circa* 1580. The equation as given here was originally propounded by Newcomb in his *Principles of Political Economy*, 1886, but was popularised by Irving Fisher, *The Purchasing Power of Money*, 1911.

houses, it is better described as the total of debits to bank accounts.

But all payments are not made by means of cheques, and account must be taken of the payments of cash. The quantity of currency in circulation and the total of bank deposits are known, and so is the total of debits to bank accounts, so that it is useful to split this item in two and to write the equation :

$$M.V + M'.V' = P.T$$

where M represents the quantity of cash in circulation, V the number of times per annum a unit of it changes hands, and $M'.V'$ represents the total of debits to bank accounts.

Velocity of Circulation

It is the practice to call V in the quantity equation the velocity of circulation of money, and we have now split it into the velocity of circulation of common money, which we have still called V, and V', the velocity of circulation of bank deposits. It is possible to compute V' by dividing the total of bank debits for a period by the average volume of deposits over that period. There is of course no method of computing the number of times a unit of cash changes hands. These quantities are commonly regarded as dependent on customs and habits of particular countries. The velocity of circulation, for instance, will depend on the proportion of his assets an individual chooses to carry about him in notes and coin. This will vary from individual to individual in the same community, but the averages for different communities may show even greater divergences. In turn there will be differences in relationship of personal assets to current income, so that the relationship between cash holding and income will depend on two variables.

When considering bank deposits we are again not concerned with a homogeneous quantity. Bank deposits may be either in personal accounts or in business accounts, and the proportions do not remain constant, nor are changes in the two necessarily in the same direction. The conditions which determine the rate of turnover of these two types of deposit are quite independent.

The personal deposits can in turn be divided into income deposits and savings deposits. The size of income deposits will bear some relation to the size of income; it is convenient to have

a reserve of bank money to harmonise the periodic inflow and the regular outflow of payments for consumption expenditure, and to meet minor emergencies. Since there is probably a tendency for people in general to maintain accustomed standards of expenditure, it is probable that when prices fall people will at first continue to purchase the same quantities of consumption goods as before, and so their income deposits will tend to increase. Similarly, they will at first tend to try to maintain consumption in the face of rising expenditure, and so run down their income deposits. Even if the money value of income deposits should remain the same, the value in terms of goods will increase when prices fall and fall when prices rise.

The savings deposits are held to provide for large irregular payments, the purchase of a new motor-car or new furniture, or even to pay for an annual holiday. Then there will be the object of providing a reserve against major emergencies, protracted illness or unemployment. Furthermore, such deposits will house funds in process of accumulation for investment, either awaiting the collection of a certain minimum sum, or the appearance of attractive opportunities for investment.

The behaviour of business deposits is still more puzzling. A business must have certain cash balances available to smooth out short-term differences between the flows of payments in and out, but it will not willingly hold balances greater than the level of reserves it really needs, as funds can be more profitably employed in the concern. When a firm is experiencing a falling off in the volume of trade, it cannot find employment in the purchase of materials, stores and labour services for the proceeds of the sales it is making, and its bank balances tend to rise. When business is active the firm may have an overdraft; that is, the bank has given it credit to honour cheques drawn on its account. The volume of business deposits at any given time will tend to be low compared with the total of debits on the account, and the velocity of circulation will be greater than normal.

It is not merely, then, the quantity of cash or bank money in existence which matters; the velocity of circulation, the rate at which cash or deposits are changing hands, is an equally important factor in determining the quantity of money payments or, what is the same thing, the total of transactions against money.

The obverse view of the velocity of circulation is willingness to hold money. If people decide that they want more notes in their wallets or bigger credit balances in their bank-books (that is, the demand for money increases), the velocity of circulation decreases. The convenience and precautionary motives for demanding money, we have already discussed. They depend on custom, and apart from the minor disturbances to the normal tenor of life at holiday times and the major disruption of accustomed ways of living which occurs in war, changes in such customs occur but slowly, hence the scant attention which was at one time paid to the influence of changes in the velocity of circulation.

There is, however, another motive for holding money which affects both personal and business holdings, and which is a potent cause of rapid changes in the volume of transactions. It is usually termed the speculative motive, and arises from anticipated changes in the relative values of commodities or securities on the one hand and of money on the other. If people suspect that the prices of commodities are going to fall, they will endeavour to diminish their holdings of commodities as far as possible, and increase their holdings of money, which they expect will buy more commodities in the future than they do to day. Similarly, if they believe that the prices of securities are going to fall, people will sell securities and hold money instead, and will thus assist in bringing about the fall in price they anticipate. There is only a certain amount of money available, and everybody cannot be a seller of either goods or securities, so prices finally reach a level at which the number of people who prefer not to sell is sufficient to take up all available supplies. Conversely, if a rise in prices is anticipated, people will try to get rid of money and acquire goods or securities as the case may be.

The Problem of the Price Level

When we were discussing the problem of exchange and the determination of prices, we used money as a symbol for goods in general, so that the value of money could on these lines be the purchasing power of money, but P in the right-hand side of the Fisher equation does not measure the value of money in this sense. The purchasing power of money over goods and services

would be an average of all prices determined by giving each price a weight representing the total expenditure of money on the corresponding commodity or service. In the Fisher quantity equation, P represents the average price of a "transaction," so that the number of times a commodity is represented in the average depends on the number of " transactions " in it. Thus, suppose we have two commodities, A and B, whose prices are approximately the same and the quantities of them consumed by a particular community roughly equal. Then, from the point of view of purchasing power over commodities, 1 per cent. changes in the prices of each would be equally important, but if A is sold direct to consumers by the producers and B passes through the hands of four middle men between producer and consumer, the influence of B on the Transactions Price Level will be five times that of A. Similarly, T bears no necessary relation to the total output of goods and services, so that neither P nor T is a quantity in which we are interested. T represents a medley of dealings in materials and finished commodities, land, buildings and fixed assets, services and securities; it includes, not only current output, but the turnover of old durable goods and securities, which is not a quantity that enters into our analysis in any other way.

Quantity of Money and Demand for Money to Hold

The Fisher approach to the problem of the relationship between the quantity of money and prices concentrated attention on the process of exchange of money for commodities, services, real assets and securities. There is an older tradition traceable to some of the earlier English writers which concentrates on the idea that people have a demand to hold money, and that the existing supply of money must be held by someone. Because the later refinement of these ideas was the work of Professors Marshall and Pigou, the equations which express their solution of the problem are called the Cambridge equations.

We used the idea of the demand to hold money to explain velocity of circulation. One concept is the inverse of the other. Marshall defined Demand to Hold in the following terms:

> " In every state of society there is some fraction of their income which people find it worth their while to keep in

the form of currency. . . . A large command of resources in the form of currency renders their business easy and smooth and puts them at an advantage in bargaining; but, on the other hand, it locks up in a barren form resources that might yield an income of gratification if invested, say, in extra furniture; or a money income, if invested in extra machinery or cattle."

The quantity actually held depends upon " balancing one against another the advantages of a further ready command and the disadvantages of putting more of his resources into a form in which they yield him no benefit." These were precisely the considerations on which we explained velocity of circulation, but while the Fisher method regarded the velocity of circulation as something which was not subject to variation in any important degree, so that the main factor in determining prices was the quantity of money, the Cambridge approach regards changes in willingness to hold money as a major source of disturbance of the price level. The exponents of this approach thought of the proportion of their resources measured in real terms which people would wish to hold in the form of money; demand is therefore not merely for a certain number of units of currency, but for the value of a certain fraction of total resources, and here we get the connection with the price level.

Professor Pigou introduced the simplifying assumptions that the total income of the community can be measured as a quantity of wheat R; it chooses to hold legal tender equivalent to a proportion k of this quantity; the number of units of legal tender is M, and one of these units exchanges for a quantity of wheat P.

Then the demand schedule to hold money is expressed by the equation:

$$P = \frac{kR}{M}$$

But people do not hold all their cash reserves in the form of legal tender; they also possess bank deposits, and allowing for these, the quantity of income which they will wish to hold in the form of legal tender will be kRc. Bankers will require legal tender as reserve against deposits they hold; let the proportion be h. The quantity of real income to be held as bank deposits will be :

$$kR\,(1 - c)$$

and the cash held against it will be the equivalent of h times

this quantity. The total demand for legal tender will therefore be given by:

$$P = \frac{kR}{M}[c + h(1 - c)]$$

Another equation in the same tradition was developed by J. M. Keynes[1] from the idea "that what a holder of money requires is a quantity of real-balances which bears the appropriate relationship to the quantity of real transactions on which he employs his balances."[2] The real-balances were measured in "consumption units"; that is, units "made up of a collection of specified quantities of their [the public's] standard articles of consumption or other objects of expenditure." People are thus considered to wish to hold an amount of cash capable of purchasing a certain fraction of the collection of things they are in the habit of purchasing per unit of time. Part of this sum they will want to hold in legal tender and part in bank deposits. So if n is the total quantity of cash and the public wish to hold k consumption units in the form of notes and coin and k_1 units in the form of bank deposits, r is the proportion of bankers' reserves of notes and coin to their deposits and p the price of a consumption unit; then we have:

$$n = p(k + rk_1)$$

The Cambridge equations are based on the inevitable equality of the demand to hold cash and the quantity of cash issued. The demand to hold is originally a demand to have command over some quantity of goods, services and assets; the money value of this collection is found by multiplying the number of units by a price which is designed to be a measurement of the general level of prices. They give a broad general picture of the relationship between the quantity of money and the price level, but the picture is often so lacking in detail as to be misleading.

Defects of the Quantity Equations

We can illustrate this by looking for a moment at the Fisher equation. There has been a tendency to regard price and the quantity of money as the only factors subject to variation, so that doubling the quantity of money would double prices.

[1] *Tract on Monetary Reform*, Chap. 3.
[2] *A Treatise on Money*, Chap. 14 (1).

Putting it perhaps more accurately, it was not assumed that the velocity of circulation could not change (it was dependent on habits of the people concerned and these clearly can change), but the possibility of changes in the quantity of money inducing changes in the velocity of circulation was commonly ignored.[1] Yet a doubling of the quantity of money could not be brought about instantaneously, and people anticipating that the purchasing power of money was going to fall would try to get rid of money. The velocity of circulation would increase and the rise in prices would probably be greater than 100 per cent. Similarly for a decrease in the quantity of money, people realising what was being or was about to be done would decrease their purchases and try to hoard money, in order to spend it when its purchasing power had increased. The fall in prices in such a case would be more than proportionate to the decrease in the quantity of money.

The Problem of Measuring Purchasing Power

We have seen that in the case of the Fisher equation, P measures the average price of a transaction, and its size is influenced by the number of transactions in which a given article figures. Professor Pigou, in his version of the Cambridge equation, uses the quantity of wheat a unit of money will buy, i.e. the wheat price of money, but this assumes that all other prices bear a constant relation to the price of wheat, so that the wheat price of everything except money is constant.

We are indeed facing the fact that any attempt to measure the purchasing power of money as a whole must be a very arbitrary affair. The statistical method of computing changes in the value of money is by means of an index number. Purchasing power is not an absolute quantity capable of direct measurement; it is not possible to say that the purchasing power of money today is so much, but it can be said that it is so much greater or less than it was a year ago. The statistical method of measuring changes in the value of money is by means of an index number of prices. If prices are higher than they were, the value of money has fallen, and conversely. We need, then, an average of prices for each occasion on which we make a comparison.

[1] Although Fisher himself was not guilty of this.

We select a collection of commodities and price it on a given date. Then, on any subsequent occasion when a comparison must be made, we again price the same collection and express its total price on the second date as a percentage of the total price on the first date. The series so obtained for a succession of dates we term an Index Number. The idea is simple enough, but the practical difficulties are considerable. If we could get a price for every item of output and we knew the quantity of it produced, we could get a number which reflected purchasing power over current output, but we know neither prices nor quantities. The number of commodities which remain constant in character over an appreciable time and for which regular price quotations can be obtained is so small that it is necessary to take arbitrary decisions as to how much each influences the purchasing power of money.

On the other hand, money is not used solely for the purchase of current output. It is used to pay for personal services and to purchase securities. It therefore becomes necessary to decide such questions as the relative effects on the purchasing power of money of a 5 per cent. rise in the price of wheat and a 5 per cent. rise in the prices of gold-mining shares. The answer must be entirely arbitrary, and as a result some writers have decided with Professor Pigou that it is no more arbitrary to consider that the price of wheat measures the purchasing power of money, or with Lord Keynes to use prices of consumption goods only.

The Demand for Cash Reconsidered

If the only purpose for which the members of a community wanted to hold either cash or bank deposits was in connection with the purchase of consumption goods, the Keynes equation would be a complete and accurate statement of the relation between the quantity of money and the price level. If, for instance, people wanted to hold enough cash to buy one week's supply of consumption goods and bank deposits sufficient to buy one month's supply, then the quantity of cash they wish to carry in their pockets, being the total price of one week's supplies, depends on the prices of consumption goods, taking the size of weekly consumption as being fixed. So if prices rise 10 per cent., people will need 10 per cent. more cash in hand.

Similarly, too, if people increase their consumption per week and do not change their habits in any other way, they will require more cash in order to have command over one week's consumption. Consumption per week will on an average be equal to production of consumption goods per week, so that the demand for money is related to the size of output. We have already seen that the quantity of money is also related to the price of output. Similarly, the quantity of bank deposits which people want to hold is, under the simplified assumption we have adopted, related to the quantity and price of output. The bankers' demand for common money is only a fraction of the volume of bank deposits. So including bankers' deposits at the Bank of England as common money, the cash required for this purpose is 8 per cent. of the value of one month's consumption. The total demand for cash under these assumptions would be approximately $1\frac{1}{3}$ times the value of one week's output of consumption goods.

Consumption, Savings and Business Deposits

Once we try to improve on the very simplified picture of reality we have been using and attempt to insert more detail, the Keynes formula becomes so complicated that its value as an expository device disappears. Indeed, this sort of difficulty eventually led its author to abandon its use, but in tracing the reasons for its shortcomings we shall discover useful material for building a fresh model.

Besides needing cash and bank deposits to enable them to maintain an even rate of consumption, people want them as savings; reserves against emergencies; provision for old age; savings for the purchase of a house or some other expensive durable consumption good; the accumulation of a "masse de manœuvre," for speculative or business purposes.

In all productive activity, money is continually being exchanged for materials, and goods are being exchanged for money, but the volumes of the two transactions per day or per week are not balanced; there are buying seasons and selling seasons, so that at certain times a business has more stock and less money and at others less stock and more money. The seasons do not synchronise for all businesses, so that at any given time the banks will be holding balances which some businesses are

waiting to employ at the appropriate time. Again, businesses must pay wages regularly every week or month, but payments for work done may come in at much longer intervals. Balances are held so that advantage can be taken of particularly favourable opportunities to invest in materials, plant, buildings or even in securities. The operators in any market, whether in materials like rubber or copper or in Bills of Exchange or in Stock Exchange securities, must have balances at their disposal which they can use to take up temporary excesses of supply over demand in their market.

There is thus some degree of separation between consumption deposits, savings deposits and business deposits, though funds may at times be attracted out of one category into another. There is also a considerable degree of separation between sub-groups inside these three categories. Balances normally employed in connection with retail trade are not readily diverted into, say, speculation on the rubber market, and funds normally employed in the house-property market will not readily find their way into, say, the finance of the manufacture of vacuum-cleaners. It is the main purpose of the British banking system to enable this transfer of funds from one market to another to take place, but that is another story. So far as the individual businessman is concerned, at times when the inflow of funds into his business is greater than the outflow, he allows his bank balance to pile up, but presumably, if there is sufficient attraction to use it elsewhere, he will do so. The form of the equation should offer some explanation of the terms on which one form of deposit can be transformed into another.

Again, we must suppose that the conditions underlying the demand to hold deposits arising from the operation of a grocery store are quite different from those influencing demand to hold deposits on the part of a jobber on the Stock Exchange; demands measured as a number of days' turnover of business will differ, so that our equation should contain a large number of terms like rk_1, each relating to some particular type of demand to hold deposits.

Once we have admitted a variety of demand conditions for deposits, we have also to take account of the different price levels. If we are concerned with the demand of Stock Exchange operators to hold cash, then the price level which is relevant to

their operations is that of securities. The total demand for money would thus become a sum of a series of products each representing some " real " demand to hold money multiplied by a price index appropriate to the kind of transaction involved.

There is thus the possibility that an increase in the quantity of money can be taken up to increase deposits of a particular type only, or, to put it in another way, the increased quantity may go to increase circulation in one market only, in which case one set of prices only would rise. So an increased quantity of money might be absorbed by the Stock Exchange to finance an increasing volume of transactions, and security prices might rise substantially while other prices remained stationary.

The results of this conclusion, that changes in the quantity of money affect different groups of prices in different ways, are far-reaching. In classical theory, as it was developed by Walras, Pareto and Edgeworth, we find a Theory of Value which provides a series of demand and supply equations whose simultaneous solution gives us all quantities of factors used, all quantities of commodities produced and consumed, and all prices in terms of money. We then require one condition to give us the quantity of money, and this they believed they possessed in the Quantity Theory. A change in the quantity of money was believed to affect all prices in similar fashion, tending to raise all together or lower all together. This we have shown to be untrue: the effect of an increase in the quantity of money depends on the manner in which it is brought about.

We are forced, therefore, to look for another method of fitting the monetary supply into our scheme of analysis. An alternative approach is suggested by our discussion of demand to hold money, and our next task will be to examine the conditions in which people are willing to increase or diminish their holdings of money.

Further Reading

Crowther, G.: *Outline of Money*, Chaps. 3 and 4.
Robertson, D. H.: *Money*, Chaps. 2 and 7.

More Advanced Reading

Greidanus, T.: *The Value of Money*, Section 1.
Keynes, J. M.: *Treatise on Money*, Books 2 and 3.

Chapter XX

THE QUANTITY OF MONEY AND THE RATE OF INTEREST

THERE are two possible ways of approaching the problem of the place of the quantity of money in the economic system. One is to watch units of money being exchanged for goods, services and securities; this is the method of the Fisher quantity equation. It concentrates attention on people's willingness to supply units of money. Wicksteed's dictum that supply is the obverse view of a demand is as applicable here as elsewhere, and leads us to consider why people should have a demand for money to hold.

We have seen that an individual may wish to hold money for a variety of reasons: to enable him to maintain a continuous flow of consumption expenditure from an income received in periodic instalments; to meet infrequent payments, which are known beforehand, but which are large in relation to a single instalment of income; to act as a reserve to meet unforeseen expenditures. Both individuals and firms will wish to hold balances for very similar reasons connected with their business operations, but instead of the objects of expenditure being meals and bus fares, rent, holidays and insurances, motor-cars, houses and surgical operations, they will be expenditures on wages and materials, replacements of equipment, extensions of scale of operations and drafts upon reserves due to failure of income to cover obligations. A trading concern must have a reserve fund on which to draw in emergency to meet its fixed costs, otherwise it will be bankrupted on the first occasion that net receipts fail to cover its fixed costs. Both individuals and firms will hold money reserves for speculative motives to enable them to seize particularly favourable opportunities to purchase goods or securities. These reserves may be held either in the form of cash or of bank deposits, according to the balance of convenience in particular instances.

The total size of the holding will bear some relation to the income of the individual in the case of consumption reserves, and will vary with the type of business and with the scale of operations in the cases of reserves held for precautionary or speculative motives. In the case of the cash reserves of a bank, as we have seen, there is in this country a conventional minimum proportion of cash to deposit liabilities, and there may be other cases where adherences to a standard is customary, but in general it will be a matter of choice, a question of balancing alternatives.

The Opportunity Cost of Holding Money

First let us be clear that this problem of the size of money holdings is entirely distinct from the problem of the quantity of savings. Saving is a matter of making expenditure on consumption less than income. The size of the money holding of an individual or a firm is the result of a decision on the proportion the money holding is to bear to the total of tangible assets and securities possessed. From tangible assets a production income of services or product is expected, from the securities a money income, and these are forgone if a cash holding is maintained. To secure homogeneity we will maintain our convention that a loan transaction always precedes an investment, even if borrower and lender are the same individual or firm. The choice before the would-be holder of cash is between the advantages of possessing money on the one hand and the money income to be derived from making a loan on the other. The opportunity cost of holding money is therefore the interest forgone.

We have thus arrived at an answer to the question, "For what is interest a payment?" *Interest is the payment for parting with the advantages of liquid control of money balances.*

Before proceeding to deal with the question of what determines the rate of interest, we must explore the relationship between the desire for liquidity and the payment for sacrificing it.

Liquidity Preference

If people had no preference for having their wealth in an accessible, i.e. liquid, form, they would be quite indifferent whether they possessed producers' goods, lands, securities or a

hoard of notes, and there would then be no charge for tying up wealth in a less accessible shape. The only circumstances in which such a state of affairs is conceivable is one where it would always be possible to find a buyer for any asset one wished to sell, instantaneously and without cost. Then indeed, all assets would be equally liquid, and clearly no charge could be made for differences in liquidity.

This state of affairs does not exist, and some formality has to be observed and some cost incurred before certain types of asset can be bought or sold. Therefore all assets are not equally acceptable in payment of debts. Assets which are more acceptable can command a premium over those which are less acceptable, and we are willing to accept limited quantities of an asset, which is entirely non-productive, because of its general acceptability as a means of payment. This kind of asset is money.

These considerations of convenience of money and costs of investment explain why the rate of interest is never likely to fall to zero, but could hardly account for a very high rate of interest.

Preference for liquid as against illiquid assets may be influenced by considerations of disadvantages attaching to one or other type of asset. Risks of loss of some kinds of asset are greater than others. Ultimately all assets are regarded as sources of consumption income, and while no asset has a constant value in terms of consumption goods, the values of all are not equally affected by the same event. We will review the main influences in turn.

Liquidity Preference and Social Security

Adam Smith long ago pointed out that rates of interest were lowest in those countries where there was the greatest degree of security to life and property and highest where there was little security. Common money is a less vulnerable form of wealth than securities in times of unrest, being more readily concealed than the real assets for which the securities themselves are symbols. Further, quite apart from the risk of destruction of the real assets which the security represents, there is the question of the sanctity of contracts. A security is in effect a loan contract, and willingness to enter into loan contracts is depen-

dent on the effectiveness with which the legal system enforces the performance of contracts. In times when the power of the State is weakened by war or insurrection, legal sanctions may be impossible to apply. In such circumstances very high rates of interest will have to be offered by would-be borrowers to induce lenders to incur the risk of making a loan.

Security of Contract

The idea of the relation between security of contract and willingness to surrender liquidity is of wider application than this. The effectiveness of legal sanctions in enforcing performance of contracts differs from one community to another according to the efficiency of the legal code; the more stringent the laws of banking, the Company Acts and the law of contract the smaller the risk of fraudulent default of a debtor.

Default, however, is not always a matter of fraud, and the credit worthiness of debtors is by no means uniform; some debtors are regarded as less credit worthy than others, because the risk of failure of the operations in which they are engaged is greater; others because they are poorer businessmen. In such cases there is a greater chance that repayment of a loan cannot be made when it is demanded or that other lenders will be unwilling to take over the loan, i.e. buy the security to which it gives rise.

At certain times the whole conduct of production seems to be fraught with more uncertainty than at others. People are afraid that prices will fall, that the demand for goods and services is declining, so that the profits of enterprise are considered likely to fall or even to turn into a loss. In such circumstances lenders incur the danger of default on the part of their creditors, are less willing to make loans and so demand a higher rate of interest.

Risks of Changes in the Value of Money

Decisions regarding investment will also be affected by anticipated changes in the value of money. If it is expected that prices will rise, so that the value of money will fall, people will be unwilling to lend money for fear that when they get it back it will be worth less than when it was lent. Thus, supposing that £100 were lent now at 5 per cent. and that when it fell due

in a year's time prices had risen so that £1 purchased only half as much as it does now, then the £105 returned would buy what could be purchased for £52 10*s*. now, quite apart from any changes in the rate of interest in the interval. Of course, it is not possible to escape the consequences of a fall in the value of money by holding a stock of cash. Still, when a drastic fall in the value of money is anticipated, people will try to sell securities for cash in order to purchase durable or storable consumption goods, and because securities are less easily turned into goods than is cash, people will be even less willing to hold investments than they will be to hold cash. The value of securities in terms of money will fall until, in periods of violently rising prices, interest rates of 100 per cent. or more for short-period loans may be experienced and long-term securities may be quite unsaleable.

Changes in the Rate of Interest

Another consideration which prevents possessors of liquid funds turning them freely into investments is to be found in changes in the rate of interest itself. The capital value of property, and therefore the quantity of consumption goods for which it could be exchanged should necessity arise, is greatly affected by the rate of interest. Thus, when the rate of interest is 5 per cent., the price of an investment with a net yield of £5 per annum in perpetuity is £100. Should the rate of interest rise to 5·25 per cent., the capital value will fall to £95·25, nearly a whole year's interest will be wiped out, and the owner will be little better off than if he had held £100 in cash. On the other hand, a fall in interest to 4·75 per cent. would raise capital value to £105·3.

Putting this in a generalised form, let a sum of money C be invested in a perpetual annuity at a rate of interest r'. The interest income will then be:

$$\frac{r'.C}{100}$$

Then:

Value of sum of money at beginning of year (invested or not) $= C$

Value of sum of money at end of year, invested $= C + \dfrac{r'.C}{100}$

Now let the rate of interest rise at beginning of year 2 to r'', where $r'' > r'$. Then the capital value of a perpetual annuity of $\frac{r'.C}{100}$ per annum when rate of interest is r''

$$= \frac{r'.C}{100} \times \frac{100}{r''} = \frac{r'}{r''}.C$$

Therefore, by his investment the lender comes to own at the beginning of year 2 a capital sum of:

$$\frac{r'}{r''}.C + \frac{r'.C}{100}$$

This investment will have paid him according to whether:

$$\frac{r'}{r''}.C + \frac{r'.C}{100} \gtrless C$$

$$\text{or } \frac{r'}{r''} + \frac{r'}{100} \gtrless 1$$

$$\text{or } 100r' + r'r'' \gtrless 100r''$$

$$\text{or } r'' - r' \gtrless \frac{r'r''}{100}$$

$$\text{i.e. } r'' - r' \gtrless \frac{r'^2}{100} \text{ approximately.}$$

Hence, if the investor expects the rate of interest to rise by more then $\frac{r'^2}{100}$ in a year, it will pay him to retain his funds in liquid form. If the rate of interest is 4 per cent., it will pay to lend at this rate, provided the rate is not expected to rise by more than 0·16 per cent. But if the rate is 2 per cent., it will pay to lend only if the rate is expected to rise by less than 0·04 per cent.

A fall in the rate of interest, on the other hand, will always bring capital gains. Consequently, willingness to lend will be affected more by this consideration when the rate is low in relation to normal levels, so that any change is likely to be upward, than when the rate is higher than normal and any change is likely to be downward.

This account of the effects of anticipated changes in the rate of interest affecting willingness to surrender liquidity has been described as leaving interest theory " hanging by its boot-

straps." However, given the possibility of changes in interest rate due to changes in liquidity preference or in costs of investment, considerations of the amount of capital gain or loss arising from this cause act as an amplifier of the amount of change in the rate of interest.

Determination of the Rate of Interest from Liquidity Preference

Having explored the relationship between willingness to hold money and the rate of interest, we can draw up a liquidity preference schedule for an individual when that individual's income measured in real terms, say in terms of wage units, remains constant. A wage unit is the price of an hour's labour. Then, if the rate of interest is low, the individual will be willing to hold a greater proportion of his assets in liquid form; that is, demand for money to hold will be greater. At higher interest rates it will be worth while to incur more inconveniences of illiquidity and to run more risks of loss, so that demand to hold money will be less and we shall get the customary inverse price/demand relationship. We can therefore represent the liquidity preference schedule by a demand curve of the usual type, sloping downward from left to right, and to it we can apply the existing quantity of money OQ and read off the rate of interest OR (Fig. 38). We have then determined the rate of interest as *the price which equates the existing quantity of money with demand to hold money balances.*

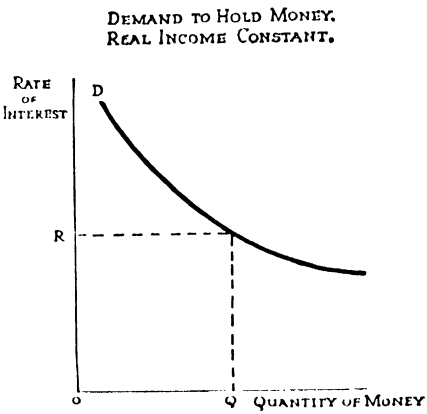


FIG. 38.

Of the formal accuracy of this analysis there can be no dispute, but the tools of economic analysis are to be judged chiefly by their usefulness. This construction was very useful to Keynes in the study of certain important problems of economic policy, but in the hands of less skilful craftsmen it is liable to

be difficult to handle. At least until we are out of our apprenticeship in economics it is advisable to use demand curves which stand still while we vary supply, and there are dangers that this one will not. A change in the quantity of money is likely to cause income, measured in real terms, to vary, and we are then concerned, not with extensions and contractions of demand to hold money, but with increases and decreases. That is, the whole curve may move to left or to right.

It is not necessary that this should take place, but there is no doubt that very often it will do so. Also it leads to the idea of *the* rate of interest as if there were only one, while it is clear that in the real world there are many rates, so that we have to seek a separate explanation for the spread of rates.

Demand and Supply of Loans

These difficulties can be avoided by adopting the alternative approach, although from the point of view of their logic both methods are equally admissible. We shall now look at the matter from the point of view of the supply and demand for loans and shall arrive at the definition—*the rate of interest is the price which equates the demand and supply of loans.* Here we shall use what we have learned of liquidity preference to explain the supply of loans, but first we must explore the conditions underlying the demand for loans. The considerations are essentially the same as those Lord Keynes used to get his schedule of willingness to invest. The demand for loans comes from two sources—from people who wish to anticipate future income and to increase consumption now at the expense of consumption in the future, and from people who wish to acquire productive assets, but who have not the necessary liquid funds for the purchase. The first element in demand is not likely to be very responsive to changes in the rate of interest, though it will vary with changes in commodity prices, in the level of taxation and the level of income of the community. People will resort to borrowing of this type, either in an endeavour to equalise consumption over time, as by anticipating an inheritance before it matures, or to meet emergencies. Changes in interest rates, within the range of normality in recent times, are not likely to affect the volume of such borrowing.

Much more important is the second type of demand. We

have already seen that, when an asset is to be purchased, the investor looks at the series of income payments he expects to get from it, and compares these with the cost of the asset; also that it is possible to express this series as a " rate of return over cost." This rate of return is not usually the same thing as the marginal productivity of capital goods at the time the investment is made, for that depends on conditions existing at the time. The rate of return will, however, be dependent on the marginal productivity of this class of capital goods at every point of time during the life of the asset, or rather on the estimates of this quantity which are made at the time of investment. On the anticipated marginal productivity of the asset during its life will depend the estimate of the gross product it will produce before it is worn out. From the gross product must be deducted the cost of the asset and also the expected cost of upkeep of the asset during its lifetime, and the net product or " yield " is then obtained.

It may be also that possession of the asset has itself particular advantages; there may, for instance, be advantages of convenience in possessing the asset rather than of hiring it each time its use is needed. To a farmer, for instance, there may be advantages in having machinery available at the moment he decides to use it rather than having its use confined to certain prearranged times. This factor is likely to be more important in relation to investments in materials, where there are advantages in having a certain stock always available for use, but in either case it is of the nature of a premium on liquidity to be added to the yield.

Further, it may be anticipated that the cost of production of similar assets may either rise or fall during the lifetime of the investment which is being made, and a profit or loss element will have to be added or subtracted to get the net return.

Schedule of Marginal Efficiency of Capital

When we have taken all these elements into account, we have the rate of return over cost or the marginal efficiency of capital. When we are considering the quantity of investment people are willing to make in a period of time, we are concerned with estimates of this marginal efficiency made in the light of the experience of the past, existing conditions and guesses about

future conditions. But it must be remembered that it is essentially an estimate of something which will be realised only over a short or long period in the future. Therefore, the estimate may change because people become more optimistic or more pessimistic without any change in the objective facts necessarily having taken place.

It is clear that different estimates of the marginal efficiency of capital will attach to different quantities of capital; the yield item, which must necessarily have a very large influence in determining the size of the final rate of return, is governed by the law of diminishing marginal products (Chapter XV). The larger the quantity of capital to be invested, the less profitable the investment opportunities which will have to be taken in order to employ it. Also an increase in the quantity of capital goods produced must eventually bring a rise in the cost of capital goods, and the yield will be reduced from this cause. It is clear then that the schedule of marginal efficiency of capital shows an inverse relation between marginal efficiency and quantity of capital employed. The price which borrowers will be willing to pay for a quantity of loans will be the marginal efficiency of the corresponding quantity of investment, so that the demand price for loans is the marginal efficiency of capital, and the demand curve for loans slopes downward from left to right in the normal fashion.

Liquidity Preference and the Supply of Loans

The factors determining the supply of loans are most conveniently dealt with in two groups, and we will exclude one group by making the assumption that there is no risk of default of payment of interest or of repayment of principal if and when it falls due. The other group of considerations which we will examine here are the set we have already discussed in relation to the problem of the demand to hold money, but now we take the obverse view of them.

Instead of thinking about the desire of individual possessors of money balances to continue to hold them, we shall think of their willingness to give them up to borrowers in exchange for the payment of interest. Thus it is clear that owners of balances will not give up the advantages of holding them without compensation. Making loans involves certain costs and

inconveniences, and for these a payment will be demanded. The greater the quantity of money balances available the less reluctant will individuals be to make reductions in those balances. Therefore we should expect that lenders will require more attractive terms for larger reductions in their holdings than for smaller reductions; that is, larger quantities of loans will need higher rates of interest to call them forth than will small quantities. The supply schedule for loans will thus show a direct relationship between the quantity of loans and the rate of interest.

Further, we have seen that the quantity of money balances is equal to the quantity of common money issued, plus the quantity of bank money created by the banking system. Given the state of willingness to make loans, the quantity which will be made at a given rate of interest depends therefore on the quantity of money. An increase in the quantity of money will increase the sums which will be available for lending at all rates of interest, and the supply curve for loans will be moved to the right for an increase in the quantity of money and to the left for a decrease.

The variability of the supply schedule will also be increased by another factor in the situation. Willingness to make loans may change at any time irrespective of changes in the quantity of money, and the lender must take these possibilities of change in one direction or the other into account. If he expects a fall in the rate of interest in the future, a loan made now will be saleable at a price higher than its original amount, when the fall in interest rate has come about. Similarly, a rise in the rate of interest will bring a fall in the capital value of the loan. We have seen that when the rate of interest is low, a very small rise in the rate will bring a capital loss equal to a year's interest income. In a given community interest rates fluctuate through time within fairly narrow limits. If the existing interest rate is near the bottom of the customary range, it is probable that any change will be in the direction of a rise, which will bring capital loss; while if the current rate is near the top of the range a change will probably be in the direction of a fall, which will bring capital gain. This effect will reinforce the tendency we have already noted for the supply of loans to be larger at higher than at lower rates of interest.

Pure Interest

When we construct a supply schedule for loans in this way and apply to it a demand schedule for loans, the point of intersection of the two schedules gives us a rate of interest at which supply and demand are equal and the loan market is in equilibrium. Such a rate is often referred to as *the* rate of interest, as if there were only one rate; or the rate of pure interest, to show that all questions of insecurity of income or principal are ruled out. The practice in the past has been to eliminate all questions of the risks attaching to individual investments from interest theory and to discuss them under the theory of profit. But eliminating all considerations of risks attaching to particular types of investment still does not leave us with a unique rate of interest. Different rates would still apply to loans for different periods of time.

Variability of Rate of Interest with Length of Loan

The picture we have built up of one common pool of loans is not an accurate one. A lender may decide that he can tie up part of his resources in a form in which it will have to remain fixed for a long period of years, but this will clearly not apply to the whole of his resources. He will wish to retain immediate control over some part which he will retain in the form of money, and other parts he will wish to keep in a form in which it can be turned into money at quite short notice. Different portions of his stock of resources will be of different degrees of liquidity.

Now, liquidity of resources may be achieved in two ways. First, the lender may make loans for varying periods of time after the fashion of the commercial banks who lend money overnight, for a week, for a fortnight, for a month, for three months and for longer periods. It is therefore possible for a bank to regain liquid control of a large part of its resources merely by ceasing to make fresh loans as old ones mature, but such a procedure requires an elaborate organisation to keep up the flow of payments and repayments. By this method a whole economy can vary its degree of liquidity, but for the individual another way is open. The assets of an individual may be quite liquid irrespective of the term for which loans have been made if the loans are readily saleable. The question of saleability,

however, raises the problem of the price at which they are saleable and the effects of changes in the rate of interest on the prices of loans for different periods are different.

The greatest fluctuations take place in the prices of perpetual loans. The price of a 4 per cent. Consol is the value at current interest rate of an income of £4 per annum for ever and that is £133 if the interest rate is 3 per cent. and £80 if the rate is 5 per cent. If the price of a 4 per cent. loan repayable in ten years' time adjusted itself so that a purchaser got 3 per cent. on his outlay, it would be equal to the present value of £4 per annum for nine years, plus £104 receivable ten years hence. If the rate of interest is 3 per cent., the value of this security is less than £133, because there is the capital loss of the difference between the present price and the £100 which will be repaid ten years hence to be taken into account. It will, in fact, be £108 5s. Similarly, if the rate of interest rises to 5 per cent., the price of the security will not fall as low as £80, because there is the prospect of £100 being repaid ten years hence.

These are the relative prices which would emerge if the yields on loans of any length were always the same, but the prices actually observed in the market do not conform to this rule.

The yield on shorter-dated loans is normally a trifle lower than that of a longer-dated loan, which suggests that investable funds do not flow equally freely into the purchase of securities of all dates, and there is no reason to expect that they should. We have already seen that some money balances are investable only for short periods. No one who had funds to invest for a few weeks would think of buying Consols except for speculative reasons. The costs of buying and selling would be large compared to the interest earned in a short period, and the risk that a price fall would bring a large capital loss is considerable. Therefore, lenders will be willing to accept lower rates for loans repayable about the time they wish to regain liquid control of their funds rather than face the possible loss on a purchase and sale of a long-dated security. The market rate of interest, say for six months' loans, will therefore depend on the demand for loans of this length and the supply of funds available for this type of loan.

The demand for six months' loans will be related to the demand for loans for somewhat longer and somewhat shorter

periods. A borrower for six months will consider the alternative possibilities of borrowing for three months and borrowing a second time to repay the first loan or of borrowing for nine months and relending the money at the end of six months when the original use for it ceases.

So too on the supply side, the rates for loans of different periods and the relative advantages and disadvantages of loans of different lengths will be borne in mind.

The Spread of Time Rates of Interest

It is clear now that what we are determining is not a single rate of interest, but a spread of rates relating to loans of different lengths. The money balances available for lending do not form a common pool which can be tapped by any borrower; rather they are split into a number of pools each available for lending under somewhat different conditions. The contents of one pool are not irrevocably committed to one type of loan, but may spill over into longer- or shorter-period loans. Thus at times when the outlook is uncertain and preference for liquidity increases, there is a tendency for the movement of supplies of loanable funds into shorter-period loans, and contrariwise, at times when a more optimistic view is being taken of the future lenders are less afraid of illiquidity and will be willing to lend for longer periods.

In times of uncertainty there is a tendency for loanable funds to seek shorter-term employment and for all interest rates to rise. In periods of expansion loanable funds seek longer-period employment and there is a tendency for all interest rates to fall. Thus at any time all changes in interest rates will be in the same direction; it will not be a case of some rates rising while others are falling, and it is for this reason that it is possible to talk of changes in *the* rate of interest, having in mind all the time a representative rate which stands for all the others. An influence which makes for higher interest rates will not affect rates for all periods of time equally, but the influence will in each case be in the direction of a rise and similarly, rates will fall together, though not to an equal extent.

The representative rate chosen by writers on interest theory has not always been the same. Those who have been concerned with general theory have chosen as their representative rate the

STAFF

long-term rate of interest which in practice they have identified with the yield to investment in the longest-term Government security available. Writers on monetary theory, on the other hand, have chosen a short-term rate, and in this country this has usually been the Bank Rate or the rate at which the Bank of England is willing to discount first-class bills of exchange for the bill market.

Changes in the Quantity of Money and Interest Rates

The rôle of the quantity of money in the loans market is now clear. The quantity of common money and bank money in existence sets a limit to the loans which can be made at any given time. Willingness to make loans determines how much of the stock of money is available as loanable funds at any given rate of interest. The process of lending is the process of redistributing money balances between those who have more than they are prepared to hold at current interest rates and those who have less than they wish to hold at these same rates.

If the quantity of money increases and willingness to lend remains the same, then those who previously had surplus money balances will now have even more than before, and those who had less money balances than they needed will now have somewhat bigger balances and their need for additional balances will be less acute.

If therefore the stock of money is to be redistributed so that at the current rate of interest everyone has no more and no less than they desire to hold, a fall in the rate of interest is necessary to extend the desire to borrow and to find a borrower for all the available loans. Lenders will readjust their holdings of money, and part of the increased supply will be taken up in an increase in their balances. They will decide that in the new situation it is not worth-while to incur the same costs and inconveniences as before to reduce their money balances, and their money holdings and their loans will both be greater at a lower rate of interest.

Cheap Money Policy

In Chapter XVIII we saw the mechanism by which the banking authority can control both the quantity of money and the rate of interest simultaneously. Now we have discovered

the reasons why such a policy is effective and the limits within which it can be effective. In the analysis above, we assumed that the schedule of willingness to lend remained the same while the quantity of money varied. If this remains true, the banking authority can increase the quantity of money and lower the rate of interest until it reaches a level at which it leaves such a small margin over the actual costs of investment that people refuse to lend any more. But it is by no means certain that the schedule of willingness to lend will remain constant. The increase in the quantity of money will have its effect on prices, and as prices rise the need for money to hold will increase, if people are to maintain a constant reserve in terms of consumption goods. Also as prices rise people may become apprehensive about the future rise in prices, and prefer to hold goods rather than money and money rather than securities. In consequence, the rate of interest may tend to rise rather than to fall, and this will come about because the supply curve for loans has moved to the left, so that a given quantity of loans will be available only at a higher rate of interest in spite of the increased quantity of money.

In these circumstances any further attempts to maintain a low level of interest by Government purchases of securities will further increase the quantity of money and raise commodity prices, thus defeating its own object.

Conclusion

Our examination of the relationship between the quantity of money and the phenomenon of interest has given us an answer to the question, "For what is interest a payment?" It is that interest is the payment for the inconveniences suffered and the risks borne when liquid funds are immobilised as loans. To the question, "How is the rate of interest determined?" we have found two alternative answers: first, that it is the rate which makes the demand for money balances equal to the existing quantity of money; second, that it is the price which equates the demand and supply of loans. If we take the first, we then want the schedule of marginal efficiency of capital to find the quantity of investment at the given rate of interest. In the second, the quantity of investment is the same as the quantity of loans, and we determine it and the rate of interest simul-

taneously. In either case, taking the *ex post* view, savings and investments are equal, so knowing investment we know savings.

Alternatively, we can take demand and supply of loans *ex ante*, and determine the quantity of investment. This will not necessarily be equal to the quantity of savings *ex ante*, since the latter is a measure of the savings people wish to make, but in the event the savings they achieve in a period may be greater or less than their original intention.

FURTHER READING

CROWTHER, G.: *Outline of Money*, Chaps. 5 and 6.
ROBERTSON, D. H.: *Money*, Chap. 5.

MORE ADVANCED READING

HICKS, J. R.: *Value and Capital*, Chaps. 11–13.
KEYNES, J. M.: *General Theory*, Chaps. 13–17.

Chapter XXI

INCOME AND EMPLOYMENT

We are now in a position to attempt to explain what determines the income of a given community and the amount of employment the production of that income will offer. Throughout our investigations we have noted the controlling power of demand over economic activity. Individuals provide productive services, because they themselves have demands for goods and services, and the use of the productive services they provide is directed by their forecasts of the community's demand for the goods and services they can produce. So Jones serves an apprenticeship as a tailor and Smith invests his capital in a worsted weaving shed, because each believes that over the period of future time in which he is interested the demand for suits will provide him with an income.

Demand alone does not, however, determine what is produced. The technical possibilities of production have to be taken into account. The quantity of every type of resource available and the efficiencies of each of them, as measured by their marginal physical products, determine the quantities of goods and services which the society can produce. We can go even farther and, while still working in "real" or physical terms, set out all the possible combinations of quantities of goods and services which the resources available to the community make it possible to produce. The technical facts of production determine the scope of choices before consumers, and demand determines which of those possible choices will be selected. But considerations of demand tell us something more than the way an income of a given size will be divided between the various goods and services on which it is possible to spend it. Considerations of demand for income in terms of effort will tell us the amount of effort the prospect of a given income will call forth, and technical considerations will tell us whether that

expenditure will produce the income which will call it forth, or whether it will produce a larger or a smaller income.

The total demand for goods and services over a period of time is measured by the earnings of that period. We saw in Chapter XVI that earnings may be looked at, alternatively, as the sum of payments to factors of production, or as the total expenditure on goods and services. The identity of these two quantities was noted more than a hundred years ago by the French economist, J. B. Say, who based on this observation his law which states that supply creates its own demand.

If the process of supply were instantaneous, Say's Law would approximate much more closely to the truth, but the supplies of goods and services available today are largely dependent on decisions to produce which were taken some time ago, and the total payments to factors of production are mainly dependent on decisions about what is to be produced in the future. There is thus no guarantee that when a thing is produced there will be any demand for it.

Earnings and Expenditures

Let us look for a moment at the definition of earnings as the total of expenditures on consumption and investment. Expenditure on consumption depends on propensity to consume. This concept is the obverse of that of time preference, which may be regarded as propensity to refrain from consumption. Habits of consumption, we have seen, are somewhat rigid, and consumption is related to income, tending to represent a larger proportion of a smaller income and a smaller proportion of a larger income. We do not get much assistance in determining income here, as the causal relationship is rather in the other direction.

This is not so with expenditure on investment goods, where the determinants are anticipations of future income and the current rate of interest. The greater the marginal efficiency of capital and the lower the rate of interest, the greater will be the expenditure on investment goods. More expenditure on goods means larger payments to factors of production for producing them and therefore the greater will be earnings. Thus it is possible for a community to increase its earnings by spon-

taneously or deliberately increasing its expenditure on investment.

Accustomed as we are to looking at the matter from the individual standpoint where it is necessary to accumulate funds by saving before expenditure on investments can take place, this statement is at first somewhat puzzling. Let us consider what would happen if the members of a community for any reason whatever showed suddenly a decreased preference for liquidity; that is, holders of money balances became willing to reduce those balances somewhat. This would have the effect of increasing the supply of loans, decreasing the rate of interest, and establishing a new equilibrium at a point lower down the scale of marginal efficiency of capital, where the demand for loans was equal to the increased supply. Investment opportunities, which were formerly beyond the margin of practicability, will now fall within it, and a demand for factors of production will arise to enable these opportunities to be taken. The earnings of these factors of production will constitute an increment to earnings and, from the increased total of earnings, there will be an increase in savings equal to the increase in investment.

This argument assumes, of course, that there are unemployed units of factor of production to be drawn into employment, or, if all the labour force is employed, that an increase in wage rates will call forth an increased amount of work per head. If there is no increased supply of factors of production in response to the increase in demand for them, there can be no increase in output. The increased volume of loans will then increase the demand for producers' goods, their prices will rise, and factors of production will be diverted from the making of consumption goods to the production of producers' goods. In consequence, the output of consumption goods will fall, and with demand increasing as wages rise, the prices of consumption goods will rise also. This in turn will increase the profitability of investment in producers' goods, and the marginal efficiency of all quantities of capital will be increased. The process known as inflation has been set in motion, and nothing but a rise in the rate charged for loans sufficient to destroy the profitability of borrowing to buy producers' goods will check its continuation until the economic system collapses.

The Volume of Employment

If we know earnings and the way in which expenditure of earnings is divided between consumption and investment, it is possible to compute the quantity of labour which will be required to produce the volumes of consumption goods and producers' goods respectively necessary to meet the existing level of demand at current prices. If Say's Law were true, then these two numbers of persons employed should equal the total labour supply, and unemployment could exist only so far as frictions prevented that movement of labour between occupations and between localities necessary to fill all vacant jobs, and so far as rigidities of wage-rates prevented the adjustment of supply and demand. Is this the whole story?

To answer this question let us return to the discussion of the theory of wages in Chapter XIV. There we saw that the demand for labour on the part of an individual firm was that quantity whose marginal physical product, when employed by the firm, was equal in value to the wage paid. For a firm in a perfectly competitive market the price of the product is constant irrespective of the quantity produced, so that the shape of the firm's demand schedule for labour depends entirely on the scale of diminishing marginal physical products. If the wage-rate is cut, the employer can increase the quantity of labour he employs until the smaller marginal product, valued at the same price as before, is equal in value to the new wage-rate. If the firm is selling in an imperfect market, we shall have to value the marginal physical product by multiplying it by the marginal revenue instead of by the price, but the same rule will hold, that a cut in wages will lead the individual firm to increase the employment it offers and a rise in wages will lead it to decrease its use of labour.

This result would probably hold true for an industry. If cotton operatives' wages were cut, the prices of cotton goods would be decreased and the demand for them would extend. More operatives would be employed at the lower wage-rate. The only possibility of this not being true is that the cut in wages should reduce the demand for the commodity, and this is not likely to happen for an individual industry.

If instead we assume a general cut in wages, what would

happen? Let us first think of a general cut, not only in wages but in payments to all factors of production in a closed community. Then we should get a cut in earnings and hence in the total monetary demand for all goods and services. Earnings, however, represent total factor costs of output, so that costs and hence prices would fall by the same amount. The most probable result would seem to be that production income would be smaller in terms of money but unchanged as regards physical dimensions and composition, and hence would require the same number of workers to make it. The net result would be that all prices of goods and services would be written down by the same proportion, and otherwise everything would go on as before. There would be no change in employment.

A cut in wages alone would have more complex effects. The wage cuts will enable prices to be reduced, and at the same time workers' incomes will be reduced in a proportion greater than the proportionate fall in prices. The wage-earners will suffer a diminution in real income and their consumption will be reduced. Recipients of other forms of income will not suffer reductions, and their real incomes will be increased as prices have fallen.

Now if this section of the community increase their consumption to absorb the goods not purchased by the wage-earners, total demand will remain unchanged, employment will be the same as before, real income will be the same as before, but will be redistributed to the advantage of non-wage-earners.

There are unfortunately reasons to believe that this is unlikely to take place. If, as is usual, non-wage-earners are at the upper end of the income scale and wage-earners at the lower end, the marginal propensity to consume of the former will be less than that of the latter, and a transfer of income in this direction will lead to a reduction in total demand and thence to a reduction in production income and in employment. The direct effects of a reduction in consumption of this type might not be very serious, but it has secondary effects, because it makes investment less attractive. Expenditure by non-wage-earners on investment is therefore likely to be reduced, and this will have further reactions on the demand for labour which will in turn react on the demand for consumption goods. A cumu-

lative tendency to decreasing effective demand and decreasing employment will have been set up, and the effect of the wage cut will be a reduction in employment and not an increase.

Wage Cuts and Employment with an Export Market

Now let us suppose the community is not a closed one but engages in international trade. The immediate effect of the reduction in the level of wages again must be to reduce the consumption of the workers concerned. The lower level of costs will enable exporters to reduce their prices for exports, and this should lead to some increase in exports. How big this increase in exports may be depends on the elasticity of demand in overseas markets for the goods of the country. If the increase in demand in response to the price fall is of sufficient size, it may absorb the goods released by the fall in consumption in the home market and more besides. In these circumstances the demand for labour will have extended as a result of the fall in wages.

The increase in employment from increased exports is not, however, an inevitable result of the decrease in wages: it is dependent on the demand for exports, and the supply of them, being sufficiently elastic. If the policy of wage cutting is attempted at a time when other countries are experiencing increasing unemployment and are raising tariff barriers against the exports of other countries, the demand for exports may not increase to the required degree.

Let us suppose that the people of Ruritania who are all dairy farmers are important buyers of British manufactures, and that they sell most of their butter to British workers. The cut in British wages will then lead to a reduction in consumption of Ruritanian butter, a reduction in incomes in that country, a reduction in Ruritanian consumption of British manufactures and a consequent decrease in employment in Britain. This is a result of the fact that the economies are so closely related that a wage cut in one has much the same effects as if the two countries were unified.

Also on the supply side, if the goods exported are of quite a different character from goods consumed at home, there will be some delay before productive capacity formerly employed in catering for the home market is converted to production for the

export market. On the degree of ease with which this conversion can take place will depend the elasticity of supply of the exported goods. If by chance the work in the export industries is of a specialised character, so unlike the work in the home consumption industries that labour cannot be transferred from one set to the other without re-training, the supply of exports will be inelastic, and no compensatory increase in employment in the export industries can be looked for, at least in the short run. On the other hand, there may be some compensatory increase in the demand of the workers in the export industry for home-produced consumption goods as, the prices of these goods having fallen, they may be willing to extend their demand.

Money Income and Employment

It is probable that a reduction in the money income of a community, say by an arbitrary cut in wages, will reduce consumption and hence reduce employment; and that a further reduction in employment will take place in the producers' goods industries, because the fall in the activity of the consumption goods industries will reduce the volume of investment. This in turn will react on the volume of consumption, which will again cause a fall in investment. At each stage employment will be reduced, but, provided wages stay fixed at the new and lower level, there is no reason to suppose the movement will go on indefinitely. We have seen that people tend to consume a higher proportion of a lower income than of a larger income, and so the successive reductions in consumption and employment will be smaller and smaller, until the system is again stable at a lower level of income, wages and employment.

Similarly, an arbitrary increase in wages will have the effect of increasing demands, first for consumption goods and next for investment goods, and in each case will bring an increase in the demand for labour to make such goods. Here, however, it is necessary to bear in mind that increases of employment can come about only if there is unemployed, or at least under-employed, labour. Unless there are people to draw into employment, or who may be induced to work harder by a higher wage, no increase in income can take place, since output cannot be increased.

The Mechanism of Change

We have now traced the effects on the volumes of output and employment in turn, of stimulating the volume of investment and the volume of consumption. A similar reaction was obtained in both cases, but the two methods are not necessarily interchangeable as instruments of economic policy. The method of stimulating investment is usually preferred, because increased investment results immediately in increased employment in the country which makes it, and this is not invariably true of increased consumption. An arbitrary rise in wages or an increase in the scales of unemployment relief may in the first instance lead to increased consumption of imported food-stuffs, and increased employment may be secondary and on a diminished scale. For a closed community there would be less to choose between the two methods.

The force of this argument may be clearer when we have examined the mechanism by which these changes take place in more detail.

An increase in wage-rates, or in unemployment relief, will be divided between consumption and saving according to the propensity to consume of the people receiving the increased incomes. In the circumstances, where this problem is most likely to arise, i.e. when there is heavy unemployment, a rise in unemployment relief is likely to be spent entirely on consumption. Next comes the division of the new expenditure between the commodities available for consumption, and this depends on what is termed the "income elasticity of demand for the commodity." This is a measure of the responsiveness of the quantity purchased to changes in income. Unlike price elasticity which we were able to discuss in terms of movement up and down a demand curve, income elasticity relates to changes in the whole demand schedule, that is, to shifts in the demand curve. As a result of the increased income the demand curves for individual commodities will move to the right by varying amounts according to the degree of income elasticity.

To trace effects correctly we must think of the results on the individual firm. We will assume perfect competition where the demand for the product of a firm is represented by a horizontal straight line and an increase in demand is represented by a

raising of such a straight line. The effect on output and hence on the employment offered by the firm depends on the shape of its marginal cost curve. If this curve is of a narrow U shape, a given increase in demand, represented by raising the horizontal demand curve a given height, will cause only a small increase in output before marginal cost is once more equal to price. If, however, the marginal cost curve is of a broad flat-bottomed type, there may be a considerable increase in output before marginal cost has risen by the required amount, and so the consequent increase in output, and hence in employment, will be greater in this case than in the previous one. A great deal of importance therefore attaches to the commodities consumers choose for increased consumption. If they select ones whose marginal costs rise rapidly, prices will rise with little increase in output, and there will, in the first instance, be little increase in employment.

The increased prices will bring increased incomes to the owners of the other factors of production, and what happens to employment at second remove depends on what these people do with their increased incomes. It is probable that their marginal propensity to consume will be less than that of the unemployed, and they will try to increase their money balances, thus stultifying the attempt to increase employment.

Apart from the shape of the marginal cost curves, it is also a matter of importance, when considering such a policy, to know how fully occupied firms may be. Eventually all marginal cost curves rise vertically when a firm has reached maximum total output, so that, if the existing output approaches the maximum, any stimulation of demand will lead to rising prices without increased employment, and thus to inflation.

If the stimulus takes the form of a lowering of interest rates, the demand for producers' goods will extend, as it will be possible for investors to move farther down their scales of diminishing marginal efficiency of capital. If the stimulus to increased investment comes from increased consumption, then the schedule of marginal efficiency of capital will be increased, and the demand curve for loans will move to the right. The quantity of loans demanded, when defined in the manner we have adopted, is also the total demand for producers' goods and for factors of production to make these goods. The demand

for labour therefore will depend partly on the total demand for producers' goods and partly on the proportions in which factors of production are being combined to make producers' goods.

What we said above about the importance of the relation between marginal cost and output is relevant here. Indeed, "bottlenecks," or cases of output being restricted by limitations on the quantity of productive capacity, which cannot be readily removed, are more important with producers' goods. The expansion of the output of a whole nation can easily be retarded by inability to increase output of an essential material or piece of equipment. We have seen in this country the limitation on output due to the shortage of coal, the supply of which in contemporary conditions was very inelastic. During the war the existence in this country of a single plant capable of making highly accurate steel tube would have greatly simplified the production of bearings for tanks and aeroplanes, and lack of equipment to make forgings and jigs and tools was a factor which retarded the invasion of France.

The occurrence of a bottleneck in a production system can therefore lead to limitations of output, so that any further stimulation of demand will cause rising prices without increased output, and will, if persisted in, lead to inflation. The ultimate limit is set when all labour is occupied, and no further increase in wages can increase the number employed or induce people to work harder. The existence of bottlenecks may cause marginal cost curves to rise vertically before all unemployment has disappeared.

Full Employment Policies

The heavy unemployment which beset almost every country in the world in the early 1930's, led to intensified study of the problem of designing economic policy to control the volume of employment. Now, the idea that governments should alleviate unemployment by embarking on constructional schemes, when the volume of employment offered by private enterprise fell seriously short of the number seeking work, was by no means a new one. The "Public Works Remedy" for unemployment was advocated as a means of alleviating mass unemployment in the later nineteenth century, but such measures were regarded as being of the nature of palliatives. Periodic occurrences of

heavy unemployment were regarded as inevitable, and public works schemes as a remedy, which would relieve the symptoms but leave the causes of the unemployment untouched.

The new element in the work of Keynes and Beveridge in the 1930's was that they believed the quantity of employment could be controlled and unemployment prevented. The policies advocated were of two types:

(1) Continuous control of the volume of employment through monetary policy.
(2) Measures to deal with emergencies when unexpected increases in unemployment occur.

The first objective was to be achieved through control of investment. Government policy should be directed towards inducing businessmen to undertake a sufficient volume of investment and thereby to offer a sufficient volume of employment to absorb all who are seeking jobs.

We have already seen that an increase in employment and the attendant increase in income can be brought about, either by stimulating consumption or by stimulating investment, but that the most direct method of achieving this end is by an increase in investment. We have furthermore seen that the volume of investment depends on the rate of interest; the lower the rate, the farther down the scale of diminishing marginal efficiency of capital it is possible to go and the greater the number of investment opportunities which can be undertaken. Each additional investment opportunity which is taken up means an increase in the demand for factors of production, including labour. Lord Keynes therefore advocated that the rate of interest should be kept sufficiently low to stimulate investment to the degree necessary to create a demand for labour which would absorb the existing supply.

Control of the rate of interest, as we have seen, lies in the hands of the monetary authority which controls it at the same time as it controls the quantity of money. By increasing the quantity of money the willingness of people to make loans is increased, so that they make available the larger quantity of loans necessary to enable the increase in investment to take place.

Willingness of businessmen to invest depends, not only on the rate of interest, but also on their confidence in the future profit-

ability of the investments they contemplate so that there may be times when the prospects for the future are so uncertain that private investors tend to hold back and wait for the situation to clarify itself. The State can then intervene and undertake investment, raising the necessary funds by means of loans.

Because of State expenditure, the earnings of factors of production will be greater than they would have been if that expenditure had not taken place. There will be an initial increment to earnings equal to the amount of the Government investment, but the matter does not end there. The owners of factors who receive this increased income will spend more on consumption than they otherwise would do, and so the incomes of the factors of production making consumption goods will be increased, indirectly, by the Government investment. They in turn will spend more in both consumption goods and producers' goods, and so the process will continue.

The amount of the increased incomes which people spend on consumption will depend on their marginal propensity to consume, and the more they consume the greater the employment they will give. The increased consumption will increase willingness to invest in two ways. It will lead to increased investment to provide for the increased consumption, and this will take the form of increased stocks of raw and semi-finished materials and also of replacements of worn-out machinery. Secondly, the increased investment may induce businessmen to revise their estimates of future levels of consumption and hence of future levels of business activity. Their estimate of the future profitability of their investments will be increased and the schedule of marginal efficiencies of capital will rise. The total effect on employment will therefore consist of the direct effect of the Government expenditure on investment, the extra employment induced by the consumption of those drawn into employment by Government expenditure and the employment created by the increase in private investment resulting from the first and second effects.

The method of lowering the rate of interest to stimulate private investment and, through investment, to increase employment, and the method of increasing employment through Government investment, are thus seen to be alternatives. Their effects are the same and the mechanism through which

those effects are attained is substantially the same, but the difference in the method of applying the stimulus is important.

If willingness to invest is at a very low level (that is, the schedule of marginal efficiency of capital has shrunk because businessmen's estimates of future profits are low), then in these circumstances a fall in the rate of interest will have little effect. If businessmen as a whole suddenly began to invest more, the investment would be profitable for everybody, but the prospect for any firm doing it alone is one of loss. The Government, however, has a direct interest in the size of the national income, for on the size of that income Government revenue from taxation depends. Therefore, it would pay the Government to employ men to dig holes and fill them up again, or to build battleships if it can thereby secure an increase in the national income greater than its expenditure on employment-giving-works. The men who are employed on Government schemes will have to be supported anyway, so that it is a matter of the choice of paying them a dole out of a smaller and decreasing national income or paying them wages out of a larger and increasing national income.

It is not of course necessary that the projects on which they are to be employed shall be useless, because there are always numbers of useful projects which could be undertaken. The essential point is that it is not necessary that the projects undertaken shall offer a rate of return over cost equal to the current rate of interest, if there are not enough projects which can pay this rate to employ the available factors of production. The individual firm cannot undertake schemes which have no prospect of paying their way. The State can, because it can set all increases in communal income, irrespective of to whom they accrue, against the extra communal expenditure, while the firm is concerned only with the share of the increased income which accrues to it.

Marginal Propensity to Consume and the Multiplier

For those who desire a more precise statement of this relationship between income, investment and employment, the following approximation may be useful. To enable a simple statement to be made, two assumptions are necessary. First, we must get over the fact that labour is not homogeneous by

thinking in terms of standard units of labour. We will assume that differences in efficiencies of labour are reflected in wage-rates, so that if one individual gets twice the wage of another, he counts as twice as many labour units. A labour unit is the amount of labour which can be bought for the wage unit, which in turn is equal to the total wage bill divided by the total number employed. It is therefore equal to average labour earnings per employed person. We shall measure the earnings of the community and the expenditures on consumption and investment in terms of these wage units.

Further, we shall assume that a given increase in investment leads to a proportionate increase in employment. This is only approximately true, but the attempt to get a closer degree of approximation will hopelessly complicate the problem.

Let us suppose that in an economy suffering from unemployment where idle equipment and stocks of materials also exist the Government draws on idle stocks of money and invests a sum ΔI in the wages of men who build roads which otherwise would not be constructed. The recipients of this expenditure, being better off than when unemployed, increase their consumption expenditure by, say, $\frac{4}{5}$ of the increase in income. This expenditure, being additional to previous consumption expenditure, increases the communal income by an equivalent sum and again the recipients of the extra income increase their consumption by $\frac{4}{5}$ of the increase, that is by $(\frac{4}{5})^2.\Delta I$. The process continues until the increments in income become so small as to be negligible, and we can then sum up the total increase in income which has taken place. The amounts form the terms of a geometrical progression, and the increase in income is given by,

$$\Delta E = \Delta I \left(1 + \tfrac{4}{5} + (\tfrac{4}{5})^2 + (\tfrac{4}{5})^3 + \ldots\ldots + (\tfrac{4}{5})^n\right) = 5\Delta I$$

The fraction, $\frac{4}{5}$ in this instance, of an increment of income consumed is termed the Marginal Propensity to Consume, and since an increment of income must be either consumed or saved, the fraction of income saved, or Marginal Propensity to Save, is $\frac{1}{5}$. Thus an increment in investment and the consequent increment in income when the reaction is fully worked out are seen to be connected by a multiplier which is the reciprocal of the marginal propensity to save, so that the smaller the proportion of an increment of income saved the larger the multiplier.

The multiplier mechanism is a "feed-back" system with a leak, the leakage being the diversion of income into savings.

The more general proof of the proposition given by Keynes runs as follows. Measuring all quantities as before in wage units and income being either consumed or invested, we have

$$E = C + I$$

Marginal Propensity to Consume is $\frac{\Delta C}{\Delta E} = 1 - \frac{1}{k}$ where $\frac{1}{k}$ is the Marginal Propensity to Save.

Since $E = C + I$, a small increment in $E = \Delta E = \Delta C + \Delta I$

$$\therefore 1 = \frac{\Delta C}{\Delta E} + \frac{\Delta I}{\Delta E} = 1 - \frac{1}{k} + \frac{\Delta I}{\Delta E}$$

$$O = -\frac{1}{k} + \frac{\Delta I}{\Delta E} \quad \therefore \frac{1}{k} = \frac{\Delta I}{\Delta E} \qquad \Delta E = k\Delta I$$

Strictly speaking k is the employment multiplier, but provided we can assume that increases in investment are approximately proportional to increases in employment the two multipliers will not differ significantly. The argument assumes that propensity to consume (or alternatively to save) remains constant, and furthermore it is assumed that the whole process is instantaneous, whereas it is more probable that a time lag will intervene between one stage of the process and the next. The argument can be amended to take account of these factors, but the essential truth of this simplified version of the process is not seriously impaired.

The multiplier mechanism requires reserves of unemployed labour, equipment and materials for its operation, and so it would appear that although it may work itself out before full employment has been reached, it cannot go on working past that point. In so far as the process operates in physical terms this must of course be so, but investment in money terms can continue after full employment has been reached, when it will result not in an increase in physical investment and income, but in an increase in the money measures of those quantities through an increase in prices.

Further Reading

MEADE, J. E.: *Economic Analysis and Economic Policy.*
MORGAN, E. J.: *Conquest of Employment.*
ROBINSON, J.: *Introduction to the Theory of Employment.*

Chapter XXII

INCOME FROM INTERNATIONAL TRADE

UNTIL quite recent years it was customary for works on international trade to be written as if its problems required for their explanation a theory quite separate and distinct from that used for the problems of internal trade, but we shall find that the same apparatus of thought serves us in both cases. The nature of the problems is unchanged, but a new emphasis has to be placed on certain features of them. Thus we have become acquainted with the effects of lack of mobility of factors of production between firms, between occupations and between localities, but factors tend to become even less mobile when national frontiers have to be crossed.

We have discussed the division of labour between individuals and between firms; now we shall be concerned with division of labour between countries, and an aspect of this problem, which is present when members of the same community are concerned, takes on a new importance when the citizens of more than one country are affected.

When technical changes take place, the reallocation of tasks which follows may destroy the livelihood of one group and benefit another. If both groups are inside the same country, some social friction will occur; but if the change results in work crossing national frontiers, there is likely to be agitation, from those who have lost their livelihood, for legislation to exclude the products of the group who have supplanted them. Finally, there are the difficulties which arise when trade is being carried on in terms of two or more currencies instead of only one. We shall review these aspects in turn.

International Division of Labour

The proximate reason for buying the product of another country instead of the home-produced article is that the im-

ported article is cheaper. The reasons for the price difference are of the same kind as those for differences in the prices of products of firms. One country may have superior natural resources. Thus Malaya can produce tin more cheaply than Cornwall, because the Malayan deposits are such that the ore can be much more easily extracted. The qualities of natural resources must always be assessed in relation to the technique of production employed, for a change in technique may modify the balance of advantage between different natural resources. The selection flotation process has reduced the advantage of using high-grade ore as compared with low-grade ore of several metals. Then, again, one country may have superior or more plentiful capital equipment; it may have a more highly skilled labour force, or workers may be willing to do a given amount of work for less pay in one country than in the other.

It is not probable that these factors will affect all types of production equally. If a country has large mineral resources or a great deal of fertile land per head of population, it is probable that it will have a relatively small amount of capital per head. Each country will therefore tend to produce those things which require a large proportion of the factors of production that are most plentiful there. Here we have the reason why it so often pays a community to specialise on the production of a limited range of commodities and to buy from other communities the other things it wishes to consume. The country has a large quantity of a factor and the opportunity to use it in an occupation where a large proportion can be used without marginal physical product falling to a very low level. Also, if the commodity produced is exported, it will be possible to expand sales much farther for a given fall in price than would be the case if the industry were catering only for home demand.

The scales of marginal physical products of labour and land occupied in wheat growing in Canada would be unaffected by the disappearance overnight of the export demand for Canadian wheat, because these scales depend solely on the physical quantities of the various factors employed, but the scales of value of marginal products would suffer a drastic fall. The eventual result of such a change in demand would be that the quantities of all factors employed would be much reduced, equilibrium points would be much higher on the scale of

diminishing marginal physical products, and the new combination of a larger marginal physical product and a smaller price would give a marginal value product for each factor employed equal to the marginal earnings of the same factor in all the other occupations open to it.

This is merely a further example of the old maxim that the degree of division of labour is limited by the width of the market. The specialisation of a large part of the resources of a country to the production of one product is division of labour on the international scale. The increased income which such specialisation makes possible can be achieved only if there is a demand for the product when it is produced. However great the resources of soil fertility, forest, mineral wealth or water power in a country may be, they do not contribute to the wealth of the area until they are producing products which someone is willing to buy.

Use can thus be made of the special qualities of areas. Vast areas in Canada or the Argentine may be devoted to wheat, in the United States to cotton, in Australia to sheep, on the Rand to gold-mining, in Malaya to rubber-growing and tin-mining, in Brazil to coffee, and so on. We should also remember that for the same reasons South Wales has specialised on coal-mining, the Vale of Evesham on fruit-growing, and Lancashire on cotton manufacture. The same theory of location of industry explains the distribution of industry on the national as on the international scale.

The physical advantages of the area for the particular form of production are only one of the conditions necessary for this geographical division of labour. If Canada could not sell its wheat or South Africa its gold and obtain other commodities in exchange, it would be obliged to curtail that form of production and to turn the factors over to other products. International division of labour thus depends on the breadth of the international market. The more freely goods are exchanged between the nations, the more closely does the distribution of the world's resources approach that in which each unit is so employed that its contribution to world income is as great as it can be. This we saw was also the principle underlying the specialisation of human tasks.

This does not mean that every commodity will be produced

in that locality where circumstances are most favourable for it. If that were so, it would be difficult to explain why English wheatland capable of producing thirty-six bushels per acre should in the past have ceased to produce wheat, and that Canadian land capable of producing only twenty bushels per acre should have been brought into cultivation to make up the deficit. The English wheatland had a more valuable marginal product when employed in some other way than it had in wheat production, so it changed its occupation. The best acre of wheatland in the world may be worth more as a mediocre factory site than for growing wheat, so it ceases to be cultivated.

The Law of Comparative Costs

The benefits to be derived from specialisation can be illustrated by the Comparative Cost Law which was once used to account for the phenomenon of differences in commodity prices between countries. It does not offer an explanation which is acceptable in the light of modern price theory, but it affords a simple illustration of the advantages derived from international trade and is worth retaining on this ground.

In its original form, this generalisation was made under the assumptions that costs and prices can be measured in terms of hours of labour, that costs per unit are constant and that exchange is by barter. It is then quite easy to demonstrate the advantages of specialisation and international division of labour.

Let us suppose that in Ruritania ten days' labour will produce either twenty units of linen or twenty units of wheat, while in Sylvania the same quantity of labour will produce either ten units of wheat or fifteen units of linen. Ruritanian labour is more efficient in the production of both commodities than Sylvanian, but it will, nevertheless, pay Ruritania to specialise in wheat growing and to import linen. If this is done, twenty days' labour in wheat growing will provide twenty units of wheat for home consumption and up to thirty units of imported linen instead of twenty of each as would have been the case without specialisation. The arrangement benefits Sylvania also. By specialising on linen and exchanging half the produce with Ruritania, twenty days' labour produces up to fifteen units each of wheat and linen instead of ten of one and fifteen of the other, as without specialisation. These figures give the

maximum amounts by which each country may benefit, but since it will pay Ruritania to import linen if more than one unit can be obtained for a unit of wheat, and it will pay Sylvania to buy wheat so long as something less than one and a half units of linen is paid for each unit of wheat, there is a range of wheat prices, from one unit of linen to one and a half units, inside which a bargain will benefit both parties.

By varying the figures in the example it is possible to show that trade will take place, except when the ratio of the efficiency of labour in one industry to its efficiency in the other is the same in both countries. Thus, if in the above example Sylvanian labour had produced fifteen units of either commodity, one unit of wheat would have been worth one unit of linen in either country and there would have been no advantage in transporting it from one to the other. Similarly, if the outputs of a given quantity of labour had been thirty of wheat and twenty of linen in Ruritania and fifteen of wheat and ten of linen in Sylvania, the price of linen would have been one and a half units of wheat in both countries, and again exchange would have benefited nobody and would not have taken place.

Now we will introduce slightly more complicated assumptions. The two countries, we will suppose, use as money gold dollars of identical weight and fineness, so that people in one country are willing to accept the coin of the other country. Furthermore, transactions in wheat and linen are the only ones taking place between the inhabitants of the two countries, and transport costs are still ignored. Assuming as before that the product of a day's labour is two units of wheat or two units of linen in Ruritania and one of wheat or one and a half of linen in Sylvania, we can now say something about the levels of wages in the two countries.

Wage-rates in Ruritania cannot be more than double those in Sylvania. If the difference is greater than this, the price of wheat in Ruritania will be greater than the Sylvanian price. If wages in Sylvania are one dollar a day, then the price of wheat will be one dollar a unit, and if Ruritanian wages are two dollars a day, the price of wheat there will also be one dollar a unit, so that no trade will take place between the two countries. If Ruritanian wages are higher than this, it will be cheaper to buy both commodities in Sylvania; trade will there-

fore be in one direction only; dollars will be exported to pay for Sylvanian wheat and linen. As the quantity of money in Sylvania increases, prices and wages there will rise, while as the quantity of money in Ruritania decreases, prices and wages there will fall. Similarly it can also be shown that wages in Ruritania must be at least one and one-third times as great as in Sylvania, for if they were lower than this level, linen would also be cheaper in Ruritania. Dollars would now be exported to Ruritania, prices and wages there would rise, while the other country would experience a corresponding fall until, when wages in the former were one and one-third times wages in the other, prices in the two countries would be the same and trade would cease.

The ratio of wages in one country to wages in the other must therefore lie between the limits of one to one and one-third and one to two. The lower limit is set by the ratio of the productivity in the most efficient Sylvanian industry, i.e. linen, to the productivity of labour in the Ruritanian linen industry. Similarly, the upper limit is set by the productivities in the least efficient Sylvanian industry and the corresponding Ruritanian industry.

To pursue the matter a stage farther: let us suppose that wages in Ruritania are $1½ per day and in Sylvania $1 per day, the two industries employing the same grade of labour in each country.

Then we have:

In Ruritania:

10 days' labour at $1½ costs $15, produces 20 units wheat at cost of $0·75 per unit.

10 days' labour at $1½ costs $15, produces 20 units linen a cost of $0·75 per unit.

In Sylvania:

10 days' labour at $1 costs $10, produces 10 units wheat at cost of $1·00 per unit.

10 days' labour at $1 costs $10, produces 15 units of linen at cost of $0·66 per unit.

It is now clear why exchange takes place and how it is advantageous to both parties. Once we have reached the stage of dealing with money costs we can drop the assumption that labour is the only factor of production employed and think of

money cost as representing payments to a number of factors. But there are still some important limitations to the usefulness of this method of analysis. In the example above, one country would specialise on wheat and the other on linen, neither would produce any of the product it imported; yet most countries produce some quantity of many commodities they import, and the discrepancy must be explained. Failure to explain is the result of our assumption of constant costs, for when costs are constant, if the imported article is available at a price lower than the cost of any part of the home supply, it is cheaper than all the home supply and the imported article will be used exclusively. It is also necessary to extend the apparatus to deal with more than two commodities and more than two countries.

The labour costs and the money costs, with which we have been dealing here, are average costs, and further developments of the method encounter all the difficulties of finding a connection between average cost and the reserve price, which we discovered when we were discussing the supply curve. We had then to abandon the attempt to construct an explanation in terms of average costs and to use marginal costs instead, and we must do the same again.

Marginal Cost and International Price Differences

We will work through this problem of international price differences in terms of wheat, and we will assume that wheat is a perfectly homogeneous commodity and that there are no differences in the kinds and qualities of wheat grown in different countries, although in fact such differences are considerable.[1] But although wheat is the same commodity the world over, it will still not be true that its price will be the same everywhere. The price of a commodity or factor is uniform if movement from cheaper price to dearer price markets is costless. In order to move wheat, costs of transport, insurance and merchanting services have to be incurred, so that the price of wheat in Liverpool can exceed the price of wheat in Buenos Aires by the costs of movement between the two centres. A greater difference than this cannot exist for long, because movement will

[1] We also assume a free market in wheat, although such a market has not existed in this country since the Wheat Act of 1932.

wipe it out, but a smaller difference can of course persist, and while it does, wheat will not move between the two centres.

The supply of wheat in Britain is made up of the contributions of many suppliers all over the world, and the size of the contribution of each of them is that quantity whose marginal cost is equal to the British price.[1] The constituents of marginal cost will not be the same for all. For home-grown wheat the marginal cost to the farmer may be greater than marginal cost to the Canadian or Argentine grower, but the cost of handling will be greater for imported grain. Ignoring differences in quality, the marginal costs delivered to British flour mills, of wheat of all origins, would be the same, because in a perfect market a miller will not pay a higher price for British wheat than he has to pay for Canadian or Argentine, and a seller in Ottawa or Buenos Aires will not accept a price less than the British price minus the costs of moving the wheat to Britain. He will increase his output so long as his marginal cost falls below this figure.

This approach to the problem shows how it is possible that part of the supply of a commodity may be home produced and part imported. It was a flaw in comparative cost theory that if the average cost of the home-produced article was greater than the average cost of the imported article, all supplies of this commodity would be imported, and if the two average costs were equal, none would be imported. In our theory any home producer who can produce at a marginal cost equal to the market price can make a contribution to supply.

The rule of equality of marginal cost is true only so long as we are dealing with a homogeneous commodity, but it is still a statement of tendency when home and imported supplies are substitutes but not perfect substitutes for each other.

Differences in Payments to Factors

Inside a community, factors of production are usually sufficiently mobile to ensure that all firms making closely related products pay much the same prices for the factors they employ, and cases of one firm enjoying an advantage over rivals through

[1] Under contemporary conditions the marginal cost of home-grown wheat will exceed the marginal cost at British mill of imported wheat by the amount of the subsidy to the British grower.

its ability to buy one or more factors very cheaply cannot be common. In international trade there is much less mobility between factors. The rich alluvial tin deposits of Malaya cannot be moved to Cornwall, where they would earn higher rents than they do in Malaya; the climatic conditions of Jamaican banana plantations can be reproduced in British greenhouses only at enormous expense, and Calcutta jute spinners and weavers cannot move to Dundee, where they could earn much higher wages. The obstacles to movement may be such that large permanent differences between the values of marginal products can persist. For such a difference in, say, wages to be permanent, it would be necessary that the demand for labour to make the product in question should be a small part of the total labour supply, otherwise the increasing demand for labour for this industry would bid up the price of labour and reduce the wage difference between the two countries. If the supply of labour to the industry was very elastic, so that the wage level was little affected, the low-wage country would be able to undersell its rivals until it had captured the whole of the trade in the commodity. This result is of course dependent on the products of the two areas being perfect substitutes, a condition which rarely occurs in reality, but although Japanese cottons were not perfect substitutes for Lancashire cottons nor Calcutta jute goods for Dundee goods, the field of substitution was large enough for the low-wage area to make considerable inroads on the markets of the high-wage area.

In the absence of barriers to trade, an area having overwhelming advantages would drive its rivals out of production. In some cases, like those of bananas or citrus fruits, this is regarded as entirely normal. It is not impossible to grow such things in Britain, though the cost would be enormous. When a manufacture has once been established in a country, however, and conditions change giving a rival such advantages that the home producer cannot compete, there is a vested interest defending its property rights, and there is then a large measure of sympathy for the industry suffering from foreign competition and support for proposals to protect it with a tariff. Nevertheless, there is the same misdirection of resources in that factors of production are being put to uses in which they are less efficient than they are in certain alternative occupations.

The supply schedule for the national market in a commodity which enters into international trade will therefore be constructed in exactly the same way as if all the producers were situated within the same political frontier. Each producer who makes a contribution to supply will contribute that quantity whose marginal cost of putting it on that market is equal to the market price. But although one condition of perfect competition is likely to be fulfilled, in that the numbers of buyers and sellers are large, yet there is not perfect mobility of factors of production between producers in different national areas. Consequently, there is no reason why average costs in one national area should be the same as average costs in another national area contributing to the same market supply.

The Effects of National Monetary Systems

What we have said so far of the trade between national areas is equally applicable to trade between regions of the same country. Any differences between international trade and inter-regional trade have lain in differences in the degree of mobility of factors of production. The problems which arise because different national areas have different currencies are also problems of lack of mobility of money. When two districts have the same common money, this form of money is perfectly mobile between them, but if the banking system is localised, the bank moneys of the two areas will not be perfect substitutes, and so bank money will not be perfectly mobile between the areas.

With national units of money, general acceptability usually ceases at the frontier because, even if it is possible to exchange foreign money for local money, some trouble and expense is commonly involved. If a large volume of international trade is to be maintained, it is necessary to make arrangements for people who possess money of one country and who wish to pay debts in that of another to make the necessary exchange.

Markets in Foreign Exchange

International trade existed long before there was any market in foreign currencies, or, as it is usually termed in this connection, foreign exchange. In the absence of a money acceptable to both parties, it is necessary to resort to barter or to methods which closely resemble barter. A cargo would be sold in one

country and the proceeds used to buy there another cargo which would be sold in a second country, and the process repeated until a cargo which could be sold in the home market was acquired. Such a method imposed severe limits on the range of goods which could enter into international trade, and an opportunity was offered to individuals who were willing to buy one currency for another, knowing that they would always be able to find somewhere a buyer who wanted that currency to settle transactions. Through a large part of the history of trade, such transactions could probably be carried on only by persons who were engaged in trade in commodities, so that they could always, in case of need, bring their money home in the form of goods by sending it back to its native land and using it to buy goods there.

When transactions between two currencies became frequent, regular prices of one currency in terms of the other would be established, and such prices came to be termed rates of exchange. These rates of exchange would be determined, like any other prices, by supply and demand. If at a given time the currency of country *A* was so plentiful in country *B* that there was more than was required to meet the current needs of people there with debts to settle in country *A*, dealers in exchange would consider whether it was likely that at some not too distant time the situation might be reversed, and there might be a shortage of *A* currency in relation to the demand for it in *B*. They would then be willing to buy the *A* currency at a price sufficiently below the price they expected to rule in the future to yield them what they considered a reasonable margin of profit. This margin would be related to the interest they would lose on their money while it was tied up in *A* currency and also to the risks of their forecast of the exchange market being wrong.

There will, nevertheless, be a limit to the profit such a dealer in exchange can exact. A resident in country *B* who holds *A* currency will not accept a given price for it if he thinks he can do better by sending the money back to country *A*, using it there to buy products of that country, importing those goods into country *B* and selling them there for *B* currency. Thus, suppose that Britain buys raw cotton from Egypt and sells machinery to Egypt. The currency units of both countries

happen to be called " pounds," but their relative values are free to change. If Egyptian pounds were very cheap in London, a holder of them would consider if it was not more profitable to send them to Egypt, use them to buy cotton and bring the cotton to Britain for sale. To do this would involve certain costs, and we will suppose that cotton worth £100 in Britain costs 105 Egyptian pounds in Cairo, and that the cost of moving the cotton from one country to the other is £5, then £100 in London cannot be worth more than 110 Egyptian pounds, otherwise it will pay to import goods instead of selling the currency to a dealer.

Similarly, if machinery worth 100 British pounds in London sells for 105 Egyptian pounds in Cairo and costs 3 Egyptian pounds to transport to Cairo and sell there, the 100 British pounds cannot be worth less then 102 Egyptian pounds. If, therefore, the exchange dealers in London offer less than 102 Egyptian pounds to would-be buyers of that currency in exchange for 100 local pounds, then it will pay such people to buy goods, send them to Cairo, and secure the Egyptian currency they want in that way. Similarly, if the dealers ask sellers of Egyptian currency more than 110 Egyptian pounds per 100 local pounds, it will pay the sellers to use their funds to import Egyptian goods instead. The limits between which the price of Egyptian pounds in London can fluctuate are therefore 102 and 110 per 100 British pounds.

Purchasing Power Parity

The limits between which exchange rates can move in this example are set by the costs of moving goods in either direction. Between the limits set in this way, rates can fluctuate under the influence of supply and demand for Egyptian pounds in London. If there were no costs of transference of goods from one country to the other, then the exchange rate could not vary from the ratio of £E105 = £B100, since this is the ratio of prices of cotton in the two countries and also of machinery in the two countries. If the Egyptian pound depreciated to, say, 106, this would be equivalent to a decrease in the price of Egyptian cotton in Britain, and more of it would be bought, so the demand for Egyptian pounds would increase and their value in terms of pounds would rise. Similarly, if the Egyptian pound appre-

ciated to, say, 104, this would be equivalent to a fall in the price of British machinery in Egypt, and the increased demand for machinery would lead to an increased demand for British pounds, thus restoring the original ratio.

Such reasoning as this lies behind the traditional explanation of the relative values of national monetary units, the so-called Purchasing Power Parity Theory. In its simplest form the theory states that the parity rate of exchange between the national monetary units of two countries may be determined by pricing a given collection of goods in the two countries and equating the total prices. Thus, if a collection of goods which costs a pound in Britain costs 4 dollars in the United States, then the purchasing power parity rate is 4 dollars to the pound. This is a normal rate about which the actual exchange rate will fluctuate. If the actual rate differs by too great an amount from the parity rate in either direction, an increased demand for the goods of the country whose money is depreciating in terms of the other will rectify the balance.

Defects of Purchasing Power Parity Theory

The example of the Anglo-Egyptian exchange rate illustrates the defects in this theory which was used to explain the long-run normal relation between the monetary units of different countries. Cotton is not produced in Britain, and so the price of it there might appear to be determined by the exchange rate, rather than being a factor in determining it, as the Parity Theory suggests. Such goods play a very important part in international trade, and to a large extent countries import goods they do not themselves produce.

Also in every country there are goods whose whole supply is produced and consumed at home, being neither imported nor exported. These may be unique products of the locality, or they may be goods which cannot travel and must be produced where they are consumed. In either case they cannot be included in the bundle of goods whose price determines the exchange value of the currency according to Purchasing Power Parity Theory. The bundle must therefore consist only of goods which enter into the country's external trade, and it is possible that the bundles representative of the trade of two countries may include no common commodity.

The only goods which are of use to us for the making up of our bundles for comparative pricing are goods which are either imported or exported and also home produced. This class may be quite large for any given country, but to make comparisons of purchasing power, we need a large common range of such goods for each pair of countries. It is not necessary that the bundle shall be the same for all countries, but between each pair there must be a large representative sample of the international trade goods of both. In practice the sample is likely to be small and unrepresentative, and comparisons may therefore have little meaning.

But although the purchasing power parity explanation of exchange rates rests on dubious evidence, yet it contains the core of a more usable idea, for the relative prices of commodities in different countries exert a fundamental influence on the flows of commodities from one country to another and hence on the flows of medium of payment.

Supply and Demand Theory of Exchange

The Purchasing Power Parity Theory, like the Labour Theory of Value or the Subsistence Theory of Wages, was an attempt to explain the normal level of a price before the supply and demand schedule technique was fully developed. We shall see the process by which exchange rates, or the prices at which currencies exchange for each other are established, much more clearly if we set out the conditions controlling supply and demand of a currency, and also we shall the better appreciate the position of the Purchasing Power Parity Theory.

Let us consider first financial transactions between two countries only, both currencies being notes or coin of no intrinsic value. The total of payments which one must make to the other constitutes the demand for the currency of the latter in the international market. Thus, Ruritanians may wish to buy Sylvanian goods and will have to pay for them; Sylvanians may formerly have invested money in Ruritania, who now must pay interest on the investments; Ruritanians may now be buying securities in Sylvania; Sylvanians may have emigrated to Ruritania, and may be sending home remittances to their poorer relatives or may be returning home with their fortunes;

the two countries may have been at war, and Ruritania may have to pay an indemnity to Sylvania; Ruritanians may mistrust the stability of their banking system or fear their Government is going to confiscate private capital and endeavour to transfer their wealth to Sylvania.

Attempts by individuals or institutions to make these payments will constitute a demand for Sylvanian shekels and a supply of Ruritanian obols on the exchange market. It does not matter whether Sylvanians are willing to accept obols, or whether they insist on payment in their own shekels; in the first case they will try to sell their obols for shekels after they have received them, in the second Ruritanians will try to buy shekels with obols before they make the transfer. Similarly, there will be a supply of shekels and a demand for obols arising out of payments Sylvanians wish to make to Ruritanians. Thus, the supply of obols and demand for shekels relate to one set of transactions and the demand for obols and supply of shekels to another. In order to get equilibrium, demand and supply must equate for each currency. This means that the supply of obols and the supply of shekels must be equal in value, and the exchange rate moves to produce the equality. So, if Ruritanians have each year payments amounting in their own currency to 10 million obols to make to Sylvanians and Sylvanians want to pay Ruritanians 5 million shekels, the exchange rate will settle at 2 obols equal 1 shekel, because at this rate supply and demand for obols are both 10 millions and for shekels both are 5 millions and the market is in equilibrium.

Long-run Level of Exchange Rates

From this analysis it might appear that exchange rates could fluctuate both rapidly and widely, but in practice they rarely do so, tending rather to fluctuate around normal levels or parities as it is customary to call them. Thus, for a long period of years the value of the British pound fluctuated within narrow margins of $4·86 U.S.A., and similar parities were established for other currencies. As we shall see presently, there is commonly machinery to ensure that fluctuations are as narrow as possible, but this machinery can only apply the principles underlying the conditions of supply and demand we are discussing.

The supplies of currencies on the exchange market vary with

the rates of exchange. If the Ruritanian obol becomes worth 0·55 shekel instead of 0·5, Sylvanian goods will become cheaper in Ruritania and Ruritanians will increase their purchases of them, thus increasing the supply of obols and tending to lower their value on the international market. Thus, the variation in the exchange rate will set up forces tending to restore the old rate. Simultaneously, Ruritanian goods will become dearer for Sylvanians, who will cut down their purchases and so restrict the supply of shekels on the market.

The efficiency of the process of adjustment depends on elasticities of demand for imports and elasticities of supply of exports. If demand for imports is very elastic, a small appreciation in the external value of the currency will greatly stimulate imports, increase the quantity of the country's currency on the exchange market and thus tend to depreciate its value. Similarly, a small depreciation in the exchange value will cut down imports, reduce the quantity of currency on the exchange market and the rate will tend to appreciate. On the other hand, if demand for imports is inelastic, as with the British demand for imported food-stuffs, imports tend to be maintained in face of the higher cost due to the depreciation of the exchange value of the pound, and do not increase greatly if the pound appreciates. The stabilising influence is thus slight.

There is not therefore the automatic adjustment which the Purchasing Power Parity Theory suggests. A fall in the exchange value of a country's currency sets up a tendency to decrease imports and increase exports, but the quantitative importance of the tendency depends on the elasticities of the respective demands and supplies.

Demands and supplies may, however, be more elastic in the long run than in the short run. Given time, factors of production can be rearranged to produce goods for which there is an export demand, and consumers may acquire tastes for unfamiliar goods and create an import trade. Two important economic controversies of the second quarter of this century have centred on the speed with which such adjustments can take place. Those who in the late 1920's argued that Germany could not pay reparations did so on the ground that her economy would collapse before the necessary demands and supplies had become sufficiently elastic to stimulate the export trade

adequately, and the advocates of reparations considered that the essential adjustments could be made speedily. In Britain since 1945 estimates of ability to increase exports to the degree required to pay for imports at the 1938 level, and at the same time to provide funds for the repayment of debts, have depended on similar considerations.

Exchange Rates with More than Two Currencies

When a third and further currencies are introduced the nature of the problem is unchanged. Before exchange markets were fully developed, and since that time in periods of disrupted international trade, there have been separate markets for each pair of currencies. Such a state of affairs existed after the two world wars, but to a much greater extent after the second, when even French francs and British pounds were of little use except to buy goods in their home countries, and supplies of goods there were very inelastic.

When international trade was active and the exchange markets flourished, an Argentine seller of wheat would accept payment in British pounds, although he did not want to spend the proceeds on British goods, because he knew that sellers of goods he did require would accept pounds in payment. Because Britain was the largest participant in seaborne trade, British pounds virtually performed the rôle of an international currency.

The presence of professional dealers in exchange smoothed the operation of these markets, for they were always willing to buy a currency which was temporarily cheaper than normal and keep it until demand for it had increased; or they would buy a currency in one banking centre where there was a glut of it and sell it in a centre where there was a shortage. These international bankers were willing to take advantage of very small fluctuations in exchange rates, and consequently the fluctuations were kept within very narrow limits.

Further discussion of these arbitrage operations, however, had better be postponed until after the discussion of the mechanism of exchange regulation in the next chapter.

Terms of Trade

When we are considering the benefit a community derives from its trade with other communities it is useful to have some indicator of changes in the balance of advantage over time.

Such indicator is found in an index of the Terms of Trade.

If exchanges of commodities were the only economic transactions between communities we need only consider physical quantities of imports and exports and compute the weight of imports obtainable for a representative ton of exports. Then if the character of both imports and exports remained the same this would tell us all we want to know. However, the existence of transactions in services and of movements of capital funds means that such a comparison does not tell us anything about the ability of a country to buy the products of another by paying for them with its own.

Consequently it is more usual to examine changes in the relative prices of imports and exports. The ratio of the price of a representative ton of exports to that of a representative ton of imports is calculated for each year in which the comparison is made. One year is chosen as the base year and the ratios for other years are calculated as percentages of the base year ratio. In practice the index number of prices of exports is divided by the index number of the prices of imports calculated with the same base year. Then for a country such as Britain which is primarily an exporter of manufactures and an importer of foodstuffs and materials the index will rise when agricultural products become relatively more plentiful in the world at large than manufactures. A manufacturing country can buy a larger quantity of food and materials for a given amount of productive effort on its own part if it should wish to do so and the terms of trade are said to have turned in its favour. This is what happened in the late twenties and early thirties of this century while during and immediately after the war agricultural products became relatively scarcer than manufactures and the terms of trade turned against manufacturing countries and in favour of agricultural ones.

FURTHER READING

ELLINGER, B.: *Credit and International Trade.*
HARROD, R. F.: *International Economics.*
LEAGUE OF NATIONS: *Network of World Trade.*
WHALE, P. B.: *International Trade.*

MORE ADVANCED READING

MARSHALL, A.: *Pure Theory of International Trade.*
VINER, J.: *Studies in Theory of International Trade.*

Chapter XXIII

THE THEORY OF INTERNATIONAL EXCHANGE

SINCE purchasing power parity rates cannot be computed, they can never have formed the basis of the precisely defined par rates which in the past remained unchanged for long periods of years. The origins of these parities are to be found in the days when the principal unit of currency was a coin, whose precious metal content was of intrinsic value equal to the face value of the coin, and such coins circulated freely. Then a British sovereign contained 113 grains of standard gold, and the parity of exchange with the dollar was that number of dollars which contained 113 grains of gold, or 4·866. Parities of exchange were thus the ratios of precious metal content.

However, gold cannot be moved about the world without cost or risk of loss, and so resort was made to money substitutes, the earliest of which was the bill of exchange. This is a promise to pay a stated sum on a stated date, usually three months after the drawing of the bill. Thus, a British purchaser of American wheat would give the U.S. exporter a promise to pay which might be presented by the exporter's London agent on the due date, and payment made in pounds. Alternatively, the bill might be payable in New York, and then the British importer would have to arrange with an exchange dealer in London to provide dollars in New York in exchange for pounds in London. Whichever method is used, the essential feature of the transaction is that some third party shall be willing to have his stock of dollars in New York decreased and his stock of pounds in London increased by the transaction. His reasons for this willingness may be any one of the following:

(*a*) He has transactions in the opposite direction also and the two sets balance. The dealer always sells a currency at a slightly higher price than that at which he is simultaneously buying it, and gets his remuneration from the difference.

(*b*) The transactions do not balance over short periods of time, because, say, British importers buy wheat at one season of the year and British exporters of manufactures have a different selling season. Dealers will buy dollars when they are available in London, hoping to sell them there at a slightly higher price when payments have to be made to New York.

(*c*) International bankers may be willing to increase their balances in London at the expense of balances in New York, because interest rates are slightly higher in London.

(*d*) Larger balances in London may be welcome because some third currency is cheaper in London than in New York, and it is intended to use pounds to buy that currency.

At this point let us remind ourselves that we are dealing with two countries which use full-bodied gold coin, but we have in fact introduced a second kind of money, very like bank money, which consists of entries on the books of international bankers. This bank money was, however, immediately transformable into gold, and this fact set limits on the difference between the par rate of exchange between two currencies and the actual rate. If the difference was too great, it paid debtors in London to sidetrack the international bankers, draw sovereigns from their bank, and ship the gold to New York. Suppose it cost half a cent in insurance and freight to send a sovereign or 113 grains of gold from London to New York, then if a London exchange dealer offered less than \$4·861 for a pound, it paid to ship gold to New York instead of using his services to make the transfer. Similarly, if dealers asked more than \$4·871 for a pound in New York, gold would move the other way. These two exchange rates are termed the gold export and import points respectively.

Gold Movements and Prices

Such movements of gold tended to restore equilibrium in a world where international trade was free, for an export of gold constituted a reduction in the quantity of money, and an import of gold involved an increase in the quantity of money. In a country which was losing gold, commodity prices were therefore likely to fall, and that country became a better market to buy in and a poorer market to sell in. Exports were

stimulated and the inflow of payments increased, thus increasing the demand for the currency of that country and raising the exchange rates. Similarly, an import of gold means an increase in the quantity of money, rising prices, stimulation of imports and discouraging of exports and a fall in the exchange rates. It is to be noted that exchange rates with all other currencies will be affected, because all prices are affected, and hence all demands for goods which enter into international trade.

Bank Rate and Gold Movements

We have seen that one of the factors influencing the willingness of international bankers to hold funds in a particular centre is the rate of interest ruling there. Once an exchange rate has moved to the gold export point, a rise in interest rate will induce bankers to allow their balances in that country to increase, provided of course that they are confident they will be able to withdraw their funds whenever they need them. When a country is on the gold standard there is little fear on this score, and so in the past, when sterling was on the gold standard and London was the banking centre of the whole world, it was the practice of the Bank of England to regulate the flow of international payments by manipulating interest rates.

It might have done this by buying and selling Government securities in the manner we have described earlier (Chapter XVIII), but a more direct means of operating on the rates for short-term loans was then effective. When the Bank of England observed a heavy drain on its gold reserves, it raised the bank rate, that is, the rate at which it was prepared to rediscount bills of exchange for the bill market. To understand the effect of this, it is necessary to look briefly at the market mechanism.

Traders who receive payment for their goods in the form of a bill of exchange are not willing to wait, until the bill becomes due for payment, to receive their money in a more usable form. So they are willing to sell a bill for slightly less than its face value, and the brokers, discount houses and banks who constitute the bill market are willing to buy bills at slightly less than their face value and keep them until they mature. Investment in purchases of bills of exchange is a convenient way of employing funds which cannot be made immobile for long

periods of time, and the rate of return earned is less than could be earned by buying even the safest Stock Exchange securities. The prices paid for bills are computed by deducting interest at a rate which varies for the kind of bill for the period which has yet to elapse before the due date. This process is termed "discounting," and the rate of interest used is termed the discount rate for the kind of bill to which it relates.

The commercial banks were both considerable buyers of bills and also lenders of money to the bill market for the purpose of holding bills. Now formerly, when the clients of the commercial banks were making payments to foreigners in greater quantity than foreigners were making payments to them, part of the bankers' balances at the Bank of England would be transferred to the agents in London of the foreign creditors. If the latter found their balances in London rising to inconveniently high levels, they would use part of them to buy gold for export.

Two consequences would then result. First, the commercial banks, finding their reserves reduced, would restrict the quantity of bank money, and would hence begin to call in loans; and secondly, the Bank of England would begin to lose gold for export and would respond by raising the bank rate.

The calling in of loans by the commercial banks would force the bill market to reduce its holdings of bills, and it has been customary for the Bank of England in such circumstances to come to the rescue and rediscount the bills already discounted. If the bank rate is above the rate at which the bills have been discounted previously, the bill market makes a loss, and so it adjusts its rates in conformity with bank rate in order to reduce its losses when, as happens from time to time, it is forced to seek the assistance of the Bank. Other short-term loan rates will also be raised to keep them in line with discount rates.

Two results then followed: the international bankers would think again about exporting their funds in gold because there was a profit to be made when the exchange rate rose again above the gold export point, and also there was a higher rate of interest to be earned, because bank rate was raised sufficiently to make retention of funds profitable. This tended to check and indeed to reverse the outflow of gold. Secondly, the steps taken to reduce the quantity of loans and of bank money

reduced the demand for goods and services, lowered prices and stimulated exports, thus providing, more slowly, a long-run adjustment which perpetuated the short-term adjustment achieved by checking the withdrawal of international bankers' funds.

In the reverse process a lowering of bank rate would lead to an increase in foreign lending, an increase in the quantity of gold coin in circulation, an increase in the quantity of bank money, a rise in prices and an increase in imports. The practice of using bank rate to effect adjustments in international payments was abandoned in Britain in the early 1930's, and although its use has been resumed in the 1950's it is clear that its effects are no longer precisely the same as they were in the days when the leading mercantile nations were on the gold standard and international trade was largely financed through the medium of sterling bills. When sterling was firmly based on gold the international bankers were willing to move funds into London to earn a slightly higher rate of interest, but sterling having been depreciated more than once, the risks of another depreciation must be borne in mind. A rise in bank rate is now regarded as an indication that the government anticipates pressure on the cash reserves of the banking system and on its reserves of foreign exchange. It is a storm cone warning of rough economic weather and its efficacy depends on the extent to which business men are induced to believe that a reduction in activity is necessary to avoid losses at a later date. The higher cost of borrowing has a direct effect but the effect on anticipations is more important.

Gold Standards

The discussion of principles of exchange in terms of full-bodied gold currencies may appear somewhat archaic, but it has the advantage of bringing to light, free of complication, certain principles which are operative in more complex circumstances. The essential characteristic of the case we have been examining is that the international bank money which was transferred from banker to banker by means of bills of exchange was freely convertible into gold coin, which was acceptable as bullion almost everywhere. Where sovereigns were not acceptable as money, they were at least regarded as pieces of gold of guaranteed weight and fineness. All other forms of the gold standard

may be regarded as devices to ensure that all forms of money, other than gold coin in circulation in the country, are freely substitutable for gold. They have been adopted because of the long-run failure of gold supplies to keep up with the demand for gold for all purposes, including use in coinage. They are all designed to economise the use of gold. Had they not been adopted, the quantity of money in all gold-using countries would have increased more slowly than the demand for money arising from the increasing volume of business transactions and a long-run fall in prices would have been engendered. Something of this sort occurred when Germany set the fashion of adopting the gold standard after 1870, and prices fell until the middle 'nineties, when increasing gold supplies caught up with increasing demand.

So long as all external claims on a country are met in gold on demand, and this means meeting them in a medium which is freely convertible into gold, the rules at which we have arrived for a gold currency system will hold for the movements of exchanges. If gold does not circulate inside the country and the internal currency is not convertible into gold, then relations between the exchange rate and the level of prices within the country need not hold.

Forms of Gold Standard

The first departure from a gold coinage system is to a *Convertible Note Issue*, and part or the whole of the money in internal circulation then consists of notes which are repayable on demand. This was the system which prevailed, with minor breaks, from the Bank Act of 1833 to 1914. Notes issued either by the Bank of England or the commercial banks were promises to pay gold and were acceptable only so long as people believed the promise would be honoured. People held notes because of the greater convenience, and bankers had to hold such quantities of gold as sufficed to meet all probable demands for conversion.

During the eighteenth and early nineteenth centuries the amount of the reserve was a matter for the discretion of the banker, but the frequency of bank failures, when bankers could not find the gold to redeem their notes, led to legislation to regulate reserves, and the successive Bank Acts prescribed the

reserves of gold which must be maintained. The unimpeachable financial stability thus achieved was an important factor in the increasing prosperity of these islands in the second half of the nineteenth century, but the system was inelastic and probably secured stability of the exchanges at the expense of the stability of the internal level of prices.

In the nineteenth century, when external trade was relatively more important than it has been in the twentieth, stability of the exchanges was perhaps of prime importance, but in the present century the system has been accused of achieving stability of exchanges at the expense of stability of the volume of employment and the national income.

A further step in the economising of gold involves the use of a *Bullion Standard* such as that adopted in this country between 1925 and 1932. With this system individual notes are no longer exchangeable for gold coin, but the Central Bank will buy and sell standard gold bars at fixed prices, the buying price being slightly below the parity rate and the selling price slightly above. Because the Central Bank is willing to buy claims against native institutions and pay for them in gold, the exchange rate against any foreign banking centre cannot depreciate beyond the point at which it pays to buy gold and transport it to the foreign centre to settle a debt. Similarly, the local currency cannot appreciate beyond the point at which it pays to import gold to settle a claim. Gold points are therefore operative just as if a gold coinage or convertible note issue were in use. As the Central Bank is under obligation to maintain a gold reserve against the notes it issues and also to buy and sell gold to all comers at the statutory prices, then the quantity of money in circulation internally will be affected by the inflow and outflow of gold. The Central Bank will automatically increase the note issue with increased holdings of gold and decrease the issue as the holdings fall. So in a country losing gold, forces are set up automatically which tend to decrease imports and increase exports and so rectify the flow of payments.

The final stage of gold economy is reached when the Central Bank reserve consists, not of gold but of a currency based on gold. The reserve will consist of claims against the governments of countries which are on the gold standard and of

deposits with banks, usually the Central Banks, in gold standard countries. Such a country is then described as being on a *Gold Exchange Standard.* So long as the exchange standard currency is freely convertible into the gold standard currency at a prescribed rate, that rate will be effective as the rate of exchange, or at least within margins determined by the costs of making the conversion. At one remove, the currency will thus be convertible into gold.

With all currencies which are on the gold standard the essential condition of convertibility into gold ensures that the gold value of the currency unit cannot differ from a predetermined figure by more than the costs of conversion. That figure is determined by the law of the country, and may take the form of a prescribed standard of weight and fineness, if gold coin is in circulation, or of fixed buying and selling prices at which the Central Bank must deal in gold on demand, if the currency consists of notes. The gold standard world is therefore one in which gold is a commodity that is universally acceptable in payment of debts, and where national currencies are very close, but not quite perfect, substitutes for gold.

Purchasing Power Parity and the Gold Standard

It might appear from this that the purchasing power parity explanation of exchange rates does not apply under the gold standard, where exchange rates depend on the ratios of the gold parities of the currencies; but in fact the theory approximates more closely to reality under the gold standard than without it. Under that system gold must be able to move freely, and as it is universally acceptable, it will tend to move to those places where its purchasing power over commodities is greatest, that is, where prices are lowest. As it moves to such a country, prices there tend to rise, and so the purchasing power of gold tends to be everywhere the same. As we saw before, this tendency can relate only to goods which pass in international trade, and due account must be taken of costs of transport. Under gold standard conditions purchasing power parity is therefore merely a reflection of the fact that gold is an international money. Such commodities as move freely will therefore have the same price everywhere in terms of gold after allowance for transport costs has been made. Prices in local

currency will be gold prices multiplied by the currency price of gold.

If prices in a particular country rise and gold flows out, it is of course possible that reserves of gold may prove to be insufficient. Such an event is indeed more likely to occur because of financial transactions connected with investment than with trade in commodities. Thus in 1931 foreign bankers recalled funds they had invested in the form of short-term loans in London and which London had used to make long-term investments abroad. London could not recall the loans it had made, and so it had to repay its borrowings in gold. There was insufficient gold to meet all payments demanded, and in consequence Britain had to leave the gold standard. The condition of convertibility could not be maintained.

Managed Currencies

Under the gold standard the purchasing power of the currency is allowed to take care of itself; it is determined by the purchasing power of gold. Policy is devoted to the maintenance of convertibility, which in effect means the conservation of the gold reserve. The size of the gold reserve not only determines the possibilities of maintaining convertibility; it also determines the quantity of money in circulation, since the note issue is tied to the gold reserve by legal ties and the supply of bank money to the note issue by custom and the policy of the Central Bank. It is indeed the central feature of the gold standard game that a country losing gold shall decrease the quantity of money and lower prices, and one gaining gold shall increase the quantity of money and raise prices. The internal price level, and therefore the volume of employment, thus tended to be at the mercy of gold movements, which might have nothing to do with trade in commodities or services. So if owners of capital in some other country feared for the security of their property, they might move their funds to London, lowering prices and employment at home and increasing both here. When their panic subsided, the reverse process would take place, and on both occasions the economy of this country would be needlessly disturbed under the gold standard.

Adoption of a gold standard is a method of currency management, but the term " managed currency " is usually applied to

arrangements where there is no such automatic valve gear regulating the inflow and outflow of gold, and regulation is by day-to-day decisions of an authority constituted for the purpose.

The Central Bank or a specially constituted Exchange Control Board buys claims against foreign banks, paying for them in local currency from funds put at its disposal by the Government. If it operates after the manner of the British Exchange Control in the 1930's, it acts as a reservoir for surplus foreign exchange available in the country. When there was a surplus of claims on foreign banks over the total of foreign claims on British banks, the Control bought them, and so prevented the appreciation of the pound; that is, a fall in the prices of foreign currencies in terms of pounds. When there was a surplus of claims against British banks, the Control bought them with the foreign currency it had previously acquired, and so prevented a depreciation in the exchange value of the pound or a rise in the prices of foreign currencies in terms of pounds.

In this way a Control endeavours to maintain a steady rate of exchange instead of a fluctuating one. The Control stock of foreign exchange takes the place of the Central Bank stock of gold under the gold standard, and part of it indeed may be in gold, but a substantial part is likely to be in the form of bank balances in foreign banks. There is, however, no automatic convertibility, as under the gold standard. Consequently there are no known limits within which the exchange rate is confined. The only limits are those to which the Control is prepared to see the rate rise or fall and the Control is always free to change its mind, but in any case it does not tell the market its intentions.

Power of Control to Fix Level of Rate

The Control cannot arbitrarily fix the exchange rate at any level it sees fit, or at least it cannot hope to maintain a rate unless it chooses it with discretion. If it fixes the rate too high, imports will be relatively cheap in the country and exports relatively dear; claims against the country in respect of commodities and services will exceed its claims against others, or in technical language there will be an adverse balance on commodity account. To maintain the excessively high rate, the Control will have to use its foreign assets to buy up claims

against the country, and eventually, if it persists in its policy, those assets will be exhausted. The Control can smooth out short-run fluctuations in the supply and demand for its currency on the international market, but its behaviour is largely conditioned by that supply and demand, and the major influence determining the supply and demand for claims is the volume of demand in either direction for international trade goods.

Here again we have the very important element of truth in the Purchasing Power Parity Theory, that the supply of a currency on the exchange market depends in the long run on the country's demand for imports, and the demand for its currency on that market depends on the demand for its exports. Both these demands depend on the purchasing power of the currency over the respective kinds of goods.

Arbitrage Operations

In order that the exchange market shall work smoothly and continuously, it is necessary that there shall be operators who are always willing to buy a currency in a cheap market and sell it in a dear one. If international transactions were entirely bilateral (that is, accounts had to be cleared between each pair of countries without trade with third parties entering into it), there would be no reason why the purchasing power of the pound, say, when turned into dollars should approximate to its purchasing power when turned into francs. We might have a state of affairs where a pound was worth, say, $3 and a dollar worth 350 frs., while a pound was worth 1,000 frs. instead of 1,050 frs. as its dollar value would lead one to expect.

Where transactions can be made freely between centres, such differences cannot exist for long. Dealers in London would sell pounds for dollars in New York, resell the dollars for francs in Paris, and buy pounds in Paris, receiving £1 1*s*. 0*d*. for every pound employed. The whole transaction could be completed by telegram within a very short time; the profit is certain, so if the costs of doing the business are less than 5 per cent. of the capital employed,[1] there is a safe unlimited profit so long as the difference exists. As a result of such operations, termed arbitrage operations, the increased supply of pounds in New York would tend to push the dollar value of the pound down,

[1] Gross profit of 1*s*. in the pound in above example.

and the supply of dollars in Paris would depress the franc value of the dollar until the dollar/pound rate multiplied by the dollar/franc rate was equal to the pound/franc rate and such operations were no longer profitable, or more accurately they would stop when prices had moved to the extent necessary to reduce the gross profit until it only just covered the costs of the operation.

Such operations can of course take place only where movements of both funds and goods are free. In the post-war world there has been only one internationally acceptable currency, the dollar, gold having been very scarce. This was due to the fact that the United States and the other American countries with currencies based on the dollar were the most important sources of the goods needed by war-ravaged Europe and Asia. Pounds and francs have not been generally acceptable, because Britain and France have not been able to supply goods in sufficient volume. Thus people who before 1939 were willing to accept pounds in exchange for their goods, because they wanted to use them to buy British goods or they knew they could sell them to someone who did, have since 1945 preferred payment in their own currency or in dollars. Consequently, trade agreements, particularly those made immediately after the war, have shown a tendency to try to balance trade between each pair of countries making them. Exchange restrictions have destroyed the free market in exchange, supplies of which have been strictly rationed, and arbitrage operations have been mainly confined to the black market.

Currency Groups

Just as under the gold standard some countries linked their currencies to a gold standard currency by adopting an exchange standard, so it is possible for one country to link its currency to a managed currency. Such a group was the sterling group formed in the 1930's, consisting of the British Empire, except Canada, and certain politically associated countries. The group is able to take a large volume of financial transactions out of the exchange market, and provided the economic systems of the various members are sufficiently diverse in character, seasonal movements in individual holdings of exchange may cancel, leaving a more constant stock of foreign

exchange for the group as a whole. The operation of the Sterling Group was much misunderstood in America, where it was quite erroneously believed that its purpose was to reduce trade with the United States. When, in addition to the group system, there is exchange rationing, policy is designed to secure that the quantity of scarce currency available is used to the greatest advantage, but this resulted in a different selection of American goods being purchased, rather than in any curtailing of purchases as a whole.

International Currency Management

The degree of freedom from breakdowns which the British banking system achieved during the nineteenth century, as it concentrated its gold reserves in the Bank of England and parallel developments elsewhere, gave rise to the idea that an international central bank to hold the reserves of national central banks might smooth out problems of international exchanges and economise the use of gold, supplies of which were not keeping pace with the growth of world production. After the first world war there was a hope that the Bank of International Settlements might perform such a function, but in the end its scope was confined to reparations transactions. The slump of the early 1930's brought still more difficult international monetary problems, as nations depreciated their currencies in an endeavour to lower the external prices of their goods and so to stimulate their export trade and create employment. The consequent instability of international prices made external trade even more difficult, and the volume of it diminished still further. At the same time nations lost confidence in each other's currencies, refused to hold them, scrambled for the limited supply of gold, bid up its price, and depressed the prices of commodities where currencies were tied to gold.

Although these troubles had abated somewhat by 1939, the probability of their recurrence in the post-war world was recognised, and under the Bretton Woods Agreement an International Monetary Fund was established, " to promote exchange stability, to maintain orderly exchange arrangements among members and to avoid competitive exchange depreciation." Each member contributed a sum partly in gold and partly in its own currency, thus providing a reserve of exchange which should be

available to members who were finding their reserves being depleted by an adverse balance of payments. Each member undertook to stabilise the gold value of its currency and not to alter that value by more than 10 per cent. without the permission of the Fund. Each member fixed buying and selling prices for gold at margins on either side of the par rate, so that the system is still of a gold standard type, in that the linkage with gold is retained.

Now let us suppose that Britain experiences an excess of payments out over payments in, on account of transactions in goods, services and debts. The value of sterling in terms of other currencies will fall and the sterling price of gold will rise. British reserves of foreign exchange will fall, and when the value of gold reaches the export point, the Bank of England can buy the currencies it requires from the Fund. Again, within limits it can buy foreign currencies from the Fund for sterling. Whether it buys for gold or for sterling, the Bank of England reduces the quantity of money in circulation in Britain, thus setting up a tendency to a fall in prices which should stimulate exports and correct the balance of payments. The effect is to increase the effective reserves of exchange of each of the member countries.

If, on the other hand, a country increases its exports considerably, other countries will have to buy its currency from the Fund in order to pay for their purchases, and the stock of currency of the first country held by the Fund may be depleted. If this happens, the Fund can declare that currency to be "scarce," and it can then request the country concerned to sell its own currency to the Fund for gold or to make a loan to the Fund.

The Fund will then ration the use of the scarce currency among the members wishing to use it, and so limit the exports of the country in question. A country which threatened to upset the equilibrium of the exchanges by raising high tariffs against imports would therefore find its export trade being restricted and would consequently be deterred from such a course.

Obviously the Fund is no foolproof device for overcoming all difficulties in the market for currencies. It could have been made a more effective instrument of regulation than it is if

irrelevant questions of national prestige had not been permitted to interfere. To the extent that it does make reserves of international exchange more elastic, it avoids the use, on the one hand, of currency depreciations which disturb international trade with sudden changes in relative prices and, on the other, of violent measures of deflation designed to reduce internal prices and costs, to enable exports to be increased. To the extent that it avoids these disturbances, it avoids those disappointments of the expectations of producers which lead them to decrease investment and thus to bring about a decline in income and employment.

In the past, national monetary policies have been designed to secure advantages for the groups pursuing them. In the short run, particular countries have gained advantage at the expense of others, but such gains appear to have been temporary. Particular groups inside particular countries have probably benefited over longer periods at the expense both of other countries and of their own compatriots.

On the other hand, individual countries and groups of countries have from time to time pursued policies which have tended to nullify the effects of the division of the world into distinct currency areas and to bring about a state of affairs which approximated more closely to a unified currency system. To the extent to which such measures are successful, barriers to exchange are removed, the degree of uncertainty attaching to trading operations is decreased and volumes of production are increased.

Further Reading

Crowther, G.: *Outline of Money*, Chaps. 7–10.
Sayers, R. S.: *Modern Banking*, Chaps. 3 & 6–8.

More Advanced Reading

Haberler, G.: *Theory of International Trade.*
Hawtrey, R. G.: *Gold Standard*, 1939 ed.
Keynes, J. M.: *Treatise on Money*, Chaps. 34–38.
Ohlin, B.: *Inter-Regional and International Trade.*

Chapter XXIV

FLUCTUATIONS IN ECONOMIC ACTIVITY

FOR the past century and a quarter economic activity has been subject to violent fluctuations whose character for long defied diagnosis. Fluctuations in communal incomes and in volumes of employment were not unknown in the pre-nineteenth century world, but they were readily attributable to particular causes and the effects were confined to the area where those causes arose. A harvest failure brought unemployment and a high cost of living directly to the large part of the population employed in agriculture, and indirectly to the artisans who provided the manufactured consumption goods. When harvests were good and agriculture prosperous, farmers and landlords had spare funds for investment, and there was an increase in building activity. There was a general tendency for prosperity to wax and wane in all European countries together, as usually weather conditions are uniform over most of the Continent, and harvests were everywhere similarly affected. The fortunes of war opened and closed markets, and those dependent on them profited or lost.

After 1825 a new phenomenon appeared. A period of increasing activity with rising employment, rising prices and profits and, eventually rising wages would give place to falling employment, prices, profits and wages. Then the fall would be checked, and, after a period when these quantities were steady at a relatively low level, the rise would begin again and the process be repeated. These cycles of activity, which first appeared in England, had a certain rough regularity, which is now regarded as largely fortuitous, but was long considered to hold the clue to the nature of the cycles. The first cycles of the new type occurred in England, but about the middle of the century North America and the principal European countries were being affected, and the cycles of

the last quarter of the century were virtually world-wide in scope.

The Course of a Cycle

A typical cycle began with increasing activity in the capital market and a raising of capital for the acquisition of equipment. The industries mainly affected were the new railway industry and non-ferrous metal mining which enjoyed a considerable increase in demand owing to the general increase in constructional work. The first noticeable cycle which reached its peak in 1825 coincided with the beginning of railway construction in England, but both North America and Europe were affected by that of 1847. As new capital was raised and spent on equipment, there was increased activity in the engineering industry, which had to expand to cope with it. When countries other than Britain came to be affected, a marked increase in exports of constructional goods took place. The increase in employment given directly by the constructional work would lead to increased consumption of consumers' goods, and this growth of demand would be further stimulated by a rise in wages in other occupations as a result of the competition of increasing employment in constructional industries. If the railways and mines were being constructed abroad, there would be an increased export of manufactured consumption goods, particularly of clothing. Increased consumption at home and increased exports led to increased imports of raw materials and food-stuffs.

Then the enthusiasm for investment would begin to wane and share prices would tend to fall. Some new concern which had borrowed money from bankers, in the expectation of being able to raise funds on the capital market to repay the bank, would find itself unable to do so and would become bankrupt. New construction ceased with the end of the Stock Exchange boom, as no new projects were coming along to replace those that were finished. Labour engaged in constructional work was thrown out of employment, and the fall in the demand for constructional goods brought unemployment in the heavy engineering industries in Britain. Exports of both constructional and consumption goods fell at an early stage in the recession, but the demand for imports, particularly of food-stuffs, remained high

for some time, as there was a time-lag before a falling demand for exports was reflected in falling employment, and even then people in this country tried to maintain their improved standard of living. There was thus a more unfavourable balance of payments against this country on commodity account, and gold began to flow out.

During the nineteenth century these crises were particularly marked by bank failures. When banks were very localised in their activities, the failure of an important company to which large advances had been made might bring down the bank also, or customers knowing the bank had been hit might fear for its stability and withdraw their deposits, or demand repayment of its notes in gold. Banks which were otherwise stable might be brought down by a panic loss of confidence. The result of this loss of confidence in the banking system was a drainage of gold from the banks, including the Bank of England, into the hands of the public, and the Bank of England was on several occasions during the century forced to ask for a suspension of its liability under the Bank Acts to cash its notes in gold.

It was indeed the disruption of the banking system by these recurrent crises, which occupied most of the attention of students of this phenomenon. But although the improvement of the banking system removed the bank failures, so far as Britain was concerned, the periods of depression became longer, and the suffering caused by unemployment and a fall in the national income increased in intensity.

During the present century attention shifted to the paradoxical situation that many people in the world were impoverished because they could not sell their products, while other people were suffering privation for lack of the goods the first group were unable to sell, because in turn the things which they could produce were not wanted. So it was said there were millions of people in the world living in poverty and near starvation because each group of them could not sell things which all the others wanted, but could not afford to buy.

It was often stated by technical experts that the technical problems of production were solved, but that faulty distribution led to people starving in a world which could produce enough for all to live in comfort. This was not true, as has become only too painfully obvious during and since the second

world war, but nevertheless these critics had hit on an important aspect of the problem. Although the world, then as now, had not the capacity to produce enough food-stuffs to provide the physiological minimum standard for the whole of its population, or the capacity to provide them with adequate clothing and other necessities, yet *effective demand* for these goods frequently failed to occupy that capacity to a sufficient extent to enable it to absorb all the people who wished to be employed.

As we pointed out before, if Say's Law were true and supply created its own demand, everyone who wanted to be employed would be employed. Say's Law, however, really assumes instantaneous production, when it would be possible to match a demand at a moment of time with a willingness to produce at the same moment of time. This condition is not fulfilled in the real world ; decisions to produce have to be taken now on the basis of estimates of what demand will be in the future. The volume of goods and services becoming available for consumption today, is the result of decisions taken yesterday, last month, last year, or even a hundred years ago.

People may not look so far ahead as a hundred years, but they must, in making decisions to invest in very durable producers' goods, look at a prospective stream of income stretching some distance into the future. An investment opportunity which may offer considerable chances of profit in the immediate future, may be nevertheless unattractive, because the prospects of yield over the lifetime of the asset may not appear sufficiently attractive to justify the investment. The estimates which govern the direction of long-term investment cannot be entirely objective. They must contain numerous elements which are matters of opinion, and the optimism or pessimism of the estimator is an important factor.

If therefore we can suppose that investors should be subject to waves of optimism and pessimism, we should have gone a long way towards explaining these alternations of boom and slump. When they are optimistic they are more willing to invest and expenditure on investment goods thus increases. As we saw in Chapter XX, increased expenditure on investment goods brings about an increase in the communal income and an increase in employment, which depend respectively on the

investment and employment multipliers. Similarly, when investment falls because people become more pessimistic, income and employment fall. But what ground have we for believing that waves of optimism and pessimism should in fact occur?

Technical Change and Fluctuations

A major technical discovery, such as the introduction of steam traction in the nineteenth century, creates a swarm of new investment opportunities. The new investment opportunities offer a higher rate of return than that which is offered by existing investments. The whole scale of marginal efficiencies of capital is modified, so that more capital appears in it against each rate of return. The demand schedule for loans having increased and the supply schedule having remained the same, the quantity of investment made should increase and the rate of interest rise. It might appear that the economic system should find a new equilibrium at a higher rate of interest corresponding to the increased marginal efficiency of capital. The argument, however, presupposes that the conditions of supply of loans are known and remain fixed, but this condition does not hold.

Such a movement begins in a situation where there is unemployed labour and where productive equipment is also idle or not working at full capacity. In this state of affairs the increase in investment leads to an increase in employment and thus to larger labour incomes. This increase leads in turn to increased consumption and increased incomes for all who are making consumption goods. From the increased communal income more is spent on investment goods and employment is still further increased. The multiplier process has been set in train. Both demand and supply of loans are growing and, as long as supplies of all factors are reasonably elastic, the process can go on and the communal income will grow.

Revision of Estimates of the Future

As the profitability of investment increases, businessmen will be more optimistic in their estimates of future profits, and this optimism will take the form of reductions in the margins they allow for risk. If a certain project can yield a return of 8 per cent. and formerly they allowed a margin of 4 per cent. for risk,

this meant that they looked at the investment as one which would yield them 4 per cent. So if they demanded a return of 5 per cent. net, the investment opportunity would not be taken up. If now they considered that 2 per cent. would cover the risk factor, the opportunity would be worth taking. Indeed, it would be worth increasing investment in this direction, until through the action of diminishing returns the yield fell to 7 per cent., as after deducting 2 per cent. risk premium the return of 5 per cent. would still be obtainable.

With increasing demands, estimates of yields before deduction of risk premiums will also be revised in an upward direction but the most optimistic phase comes after the prices of consumers' goods have begun to move upward. This does not come at the beginning of the upward movement, because firms are working at a low level of capacity, and as their marginal cost curves will in general be of a flat U-shape, a substantial increase in output will be possible without any considerable price rise.

For this reason, when the price rise does begin, it will be rapid and profits will increase more rapidly than before, because everything which is sold is disposed of at a higher price level than that ruling when the costs of production were incurred. If businessmen become convinced that this state of affairs is going to continue, investment will increase greatly. Investors disregard present rates of return and gamble on a future increase in yield raising the capital value of investments. Capital appreciation rather than income becomes the object of investment, and although all know that the process cannot go on indefinitely, all are confident that they will be able to sell out and realise their capital gains before the crash comes.

The community invests an increasing part of its income, but as all factors of production are employed, total output measured in real terms cannot increase. Factors of production will be attracted away from the production of consumption goods to the production of producers' goods. The prices of factors of production will be bid up to attract them away from existing occupations, thus increasing the incomes of their possessors. The prices of consumption goods will rise more rapidly than those of producers' goods, so that the prospective stream of consumers' goods capitalised at the existing interest rate will

always exceed the cost of production of the producers' goods which gives rise to the stream. There is thus a perpetual stimulus to investment.

Transition from Boom to Depression

The process of expansion we have just described clearly cannot go on for ever, and the end may come in one of three ways. It is possible, though without the assistance of other influences not probable, that the shortage of consumers' goods would become so acute that prices would begin to rise very rapidly and people would endeavour to avoid holding money. The system would break down and the community would resort to barter. A system of acute inflation would have been reached.

Alternatively, some important production bottleneck might occur, and a rise in the price of some factor, more rapid than all others, might increase costs more rapidly than prices of products were rising. The profitability of investment would be affected, and somewhere people would revise their estimates of the future in a downward direction.

In each case the whole schedule of marginal efficiencies of capital would be decreased as the expectation of return from every investment opportunity was revised downwards. The demand curve for loans would consequently shift towards the origin.

The most probable cause of the break in the boom is that the monetary authority would be unable or unwilling to increase the supply of common money sufficiently to maintain the expanding volume of credit. To prevent further credit expansion it would raise interest rates high enough to make further investment unattractive. The demand for investment goods would fall, and the demand for labour to make them would shrink. The multiplier process would begin to work in reverse. Those who lost their jobs making producers' goods would cut consumption and so throw people making consumption goods out of work also. Producers of consumption goods would cease to demand new equipment for either expansion or replacement, throwing more workers out of emlpoyment. Every successive reduction in employment leads to a reduction in consumption and to further unemployment.

The Turn to Recovery

Unlike the upward movement where there was no brake within the system which would be automatically applied, there are at least certain checks to the downward process. Thus, when prices have fallen to abnormally low levels and interest rates are also low, new investment opportunities may show themselves perhaps in the form of demands which went unsatisfied in the previous boom. Thus, recovery from the depression of the early 'thirties came in Britain through an increase in house building. Interest rates had been high since 1914, and private housebuilding had been hampered by high interest costs. In the railway crises of the nineteenth century, railways might be built beyond the demand for them, say, in the United States, but the demand had been growing perhaps in Europe or India and another upward movement was started by this crop of investment opportunities. Then, after some years had elapsed and settlement in the U.S.A. had caught up with railway construction, more construction might begin there and provide the basis for another boom.

Before people regain the confidence to start investing again, they have to be assured that prices and interest rates are not going to go any lower. There must therefore be a period of stability at a low level of activity. This seems to be brought about by inelasticities in the wage system. If wages could be cut freely, the volume of consumption would fall without check. To reduce their losses employers try to reduce their commitments and try to cut wages and dismiss labour. They do not wish to disrupt their whole organisation if they can help it, and will retain a certain minimum staff to serve as a nucleus when times turn for the better. Also they may be afraid some of their rivals will take on some of their key workers if they themselves dismiss them. If such a condition exists, and if at the same time it is difficult to reduce wages and so cut expenditure, there will be a certain minimum level of expenditure on labour and a certain volume of consumption and employment will be maintained.

To maintain this level of stability employers will have to draw on reserves, so the stability can last only as long as the reserves last. If during this temporary stability a positive

factor making for an upturn, say a new crop of investment opportunities, comes along, recovery may begin, but if it does not, then there will be a drop to a still lower level of activity, and employment and the communal income will fall still lower.

Changes in Distribution of Income during Fluctuations

In the upward phase of a boom the stickiness of wages is likely to act to prevent them rising as rapidly as they might do if there were greater fluidity in the wage system. The main increase which takes place in the earnings of labour, at least in the early stages of the boom, comes about through an increase in the numbers employed, the disappearance of short-time and the introduction of overtime. The share of the national income going to capital thus increases and the share going to labour decreases. This statement relates of course to relative shares and not to absolute shares. The total wage bill increases, but it represents a smaller proportion of a larger total income. Until the phase of a rapid rise in prices is reached, both the money income of labour and the real income will increase absolutely, although the proportion of the communal income going to labour is falling.

Here we have a factor tending to bring the expansion movement to an end. A decreasing proportion of the communal income is going to the individuals who provide the factor labour and an increasing proportion to those who provide the factor capital. Now, on the whole, those who provide labour are poorer than those who provide capital, so that the marginal propensity to consume of the first group is likely to be greater than that of the second. Therefore, if an extra million pounds goes to capitalists as increased profits, a smaller part of it is spent on consumption and a greater part is invested than if it went to the workers as increased wages. Thus, as the boom progresses, the proportion of the communal income invested increases and the proportion consumed decreases, even though absolute consumption increases. The supply of investment opportunities is therefore used up at an increasing rate and, at the same time, demand for the consumption goods which the investment goods will ultimately produce does not keep pace with the growth in the stock of investment goods.

The situation arises therefore where the supply of consumption goods produced by the new investment goods exceeds the demand for them at the prices which the investors expected to receive for their products. So the expectations of the investors receive a setback and they revise their estimates of the future profitability of further purchases of investment goods, and the demand for such goods decreases. Receivers of profit incomes now attempt to increase their holdings of cash instead of making further investments, employment in the investment goods industries declines, and a downward movement in the volumes of income and employment has begun. Here it is to be noted we have a factor in the situation which automatically tends to bring about a deterioration in confidence and a downward revision of estimates of the future profitability of investment, such as we require for our explanation of the turn from boom to slump.

Propensity to Consume in the Depression Phase

Just as the rise in wages tended to lag behind the rise in the volume of employment and in other prices in the boom phase, so in the depression phase wages tend to fall more slowly than commodity prices and employment. The total communal income is now falling and the total wage incomes and profit incomes are both falling, but profit incomes are falling faster, so that the *proportionate* share of the communal income going to wage-earners increases. As we saw before, marginal propensity to consume is likely to be greater for wage-earners than for profit-earners, so that as the communal income shrinks, the proportion of it devoted to consumption increases.

The rigidity of this volume of consumption maintains a certain level of employment and serves to maintain the level of communal income, but its rigidity is originally based on the rigidity of wage-rates and can persist only so long as wage-rates are maintained. If, as a result of a setback to confidence, a serious attempt is made to cut wages in order to reduce costs, communal income and demands for consumption goods and for labour will decline, but once the fall has taken place the high level of marginal propensity to consume will again tend to prevent the fall continuing indefinitely.

Distribution of Income and Propensity to Consume

It is now clear that fluctuations in total income and employment have their origins in the volume of expenditure devoted to investment rather than in that devoted to consumption. If marginal propensity to consume is high, fluctuations in income and employment must follow fluctuations in consumption, but the conservatism of human consumption habits tends to prevent such fluctuations being very violent. Such fluctuations as do occur are likely to arise from external causes—war, pestilence, harvest fluctuations and the like. These, we have already noted, appear to have been the main causes of booms and slumps in the days when capital investment was a less important feature of the economy than it has been for a century and a quarter.

The level of marginal propensity to consume depends, not only on total income, but also on its distribution. Thus, propensity to consume is lower in a community where, say, half the national income is received by 10 per cent. of the community than it would be in a community of the same size, composed of similar individuals with the same total communal income, but where that income was distributed on an equalitarian basis. A more equalitarian community might therefore be expected to be less subject to fluctuations in income and employment than a less equalitarian community.

Further Reading

Baranowski, Tougan: *Crises Industrielles en Angleterre.*
Beveridge, W. H.: *Full Employment.*
Harrod, R. F.: *Trade Cycle.*
Lewis, W. A.: *Economic Survey*, 1919–39.

More Advanced Reading

Wilson, T.: *Fluctuations in Income and Employment.*

Chapter XXV

THE RÔLE OF THE STATE IN ECONOMIC AFFAIRS

In our study of the forces determining the quantity of each commodity and service which forms part of the output of a community and the sizes of the distributive shares in that output which go to groups of providers of factors of production, we have taken for granted the existence of certain demands and have assumed a certain distribution of wealth. We have regarded demands for consumption goods and personal services as being dependent on the native tastes of the members of the community, and on the existing distribution of wealth among them, and in demands so determined we have discovered the forces which control the set of the productive system.

But it is quite clear that the desires of individuals are not the sole forces at work, otherwise those with a taste for narcotic drugs, or those who wish to be able to buy alcohol at any hour of the day, or a tin of sardines after six o'clock, or half-a-dozen handkerchiefs on a Sunday, would be able to find someone willing to cater for the demand. In these and in many more serious matters, the State frustrates demands of individuals by prohibiting such transactions under penalty. In doing this it is enforcing a collective decision of the majority on every individual; for conduct which is in the short-run interest of an individual, if he alone acts in that manner, may be against the interests of all collectively and in the long run. Thus it is in the short-run interest of an individual to break a contract if terms he expected to be favourable when he made it turn out to be unfavourable. But if the practice of breaking contracts becomes general, any elaborate system of production and exchange is impossible and the benefits such a system can afford are unattainable. Thus, if a rule of conduct enables a community to satisfy demands at a higher level than is possible

without it, that rule will generally be enforced by collective decision, and the organisation which enforces such decisions is the State.

Any organisation of production requires rules, and the more complicated is the system the more elaborate are the rules. A community of hunters will have rules designed to prevent disputes as to ownership of carcasses, an agrarian community will have rules defining the scope of rights over land and water supplies. If through liability to external attack the community has become feudalised, there will be rules setting out the rights and obligations of each group in the land-controlling hierarchy vis-à-vis every other group.

When trade is practised on a considerable scale, rules to prevent fraud and to ensure performance of bargains are adopted, and when the practice of working for money wages supersedes the feudal land rights and service dues, we have the rudiments of a law of master and servant being established. Eventually the relations of associations of masters and of workers will be regulated.

When investments of capital in fixed equipment become necessary on a scale greater than individual businessmen can provide, the need arises for a law of limited liability. Without it, investors are reluctant to risk money in businesses they do not control, because of the danger of being called upon to make good the debts of the business in case of failure. Consequently, they demand high prospective rates of return to compensate for the risk. With limited liability their loss is limited to the sum ventured, and an individual may, by investing small sums in each of a large number of businesses, diminish the risk of loss of any substantial part of his fortune. Business can therefore be financed more cheaply with limited liability than without it, investment will take place on a larger scale, and the communal income will be greater.

Any form of organisation of production involves the existence of certain stable relationships between groups of individuals as producers, and it is the prime function of the State to preserve the stability of these relationships. It makes rules to ensure their preservation, and exercises the coercive power to enforce obedience to the rules it makes.

The Legal Framework of Economic Activity

We are not going to attempt a complete catalogue of all the legislation which regulates the process of production at the present time, but a partial review will reveal its character and illustrate the modifications in productive processes it involves.

The establishment of a new business unit involves immediate attention to a complicated body of law. All but the smallest concerns, in this country to-day, are companies with limited liability; that is, the liability of the proprietors is limited to the sums they have undertaken to subscribe. This means they are subject to the Company Acts, which prescribe a procedure for the establishment of a company, the information which must be given to would-be subscribers and the form in which the accounts of its subsequent operations shall be published. In erecting its premises, it must conform to local town-planning and building regulations, and for some years permits for materials have had to be obtained from the Central Government. The design, lay-out and method of operation of machinery must conform to the Factory Acts. If the product falls in the appropriate categories, it must conform to the standards of the Food and Drugs Acts, and this may affect the choice of materials and methods of production. Relations between the firm and its employees are regulated by the Law of Master and Servant, by the Workmen's Compensation Acts and by specific legislation in certain trades. Mining operations in particular are the subject of much specific regulation. The transport arrangements which the firm makes to carry away its products are subject to regulation by the Transport Acts. If part of the product is sold abroad, Exchange Regulations determine the ways in which it can utilise the foreign currency so earned. So we might continue with legislation affecting advertising and methods of sale. Finally the acquisition of the goods by the public may be subject to rationing.

All such regulation has its effects on supply and demand schedules. The enforcement of safety regulations in mines increases production costs, and supply schedules are raised. The restriction of the quantity of lead in a food-stuff to one or two parts per million may mean that lead vats and

piping cannot be used in its manufacture and something more expensive must be substituted. Again, costs will be raised.

Rationing and Price Equilibrium

In the case of goods rationed on the basis of a fixed quantity per head, the normal rule that price shall equal the marginal rate of substitution of the commodity for money does not hold. The individual is prevented from extending purchases until a further unit of commodity is, in his estimation, only just worth the price which has to be paid for it. A shilling's-worth of rationed goods is valued more highly than a shilling's-worth of unrationed goods, hence the scope for black markets. Neither is there any assurance that a shilling's-worth of rationed meat is considered equal to a shilling's-worth of rationed butter. Such equalities can only be the result of free choice of quantities purchased.

When goods are rationed on a points system, it is in effect necessary to surrender two currencies to obtain them. Again, the rule of equality of price and marginal rate of substitution breaks down, because money and points are not interchangeable. If the buyer secured equality of points values and marginal rates of substitution of points for goods, there would be no reason to expect that equality of money price and marginal rate of substitution of goods for money was secured at the same time. If high-pointed goods were also high priced, a rough equilibrium might be reached, but if high-pointed goods were cheap and low-pointed ones dear, the rule would not hold even approximately.

The idea behind the rationing is of course that it is important that all should be able to satisfy needs at some minimum level. In a free market there is allocation by price, in that those with the deepest purses can purchase large quantities before the equilibrium point is reached; while others may reach this point before a quantity regarded as adequate according to some external standard has been reached. Let us suppose a government decides that fresh meat supplies are just adequate to provide the minimum necessary for physical efficiency if shared out equally among the population. If it introduces a system of

rationing it will be able to sell at a price equal to marginal cost.[1]

Everyone might then have the minimum required for health, but many would consume less than they were willing to consume at the current price. Their marginal rate of substitution of meat for money would exceed the market price. In the absence of rationing, such persons would be able to bid up the price of meat. In normal circumstances their efforts to increase their consumption would lead to an increased production of meat. The marginal cost of meat might rise slightly, but as supplies are normally fairly elastic, the rise would not be considerable. The increased consumption of richer people would not be at the expense of decreased consumption of poorer persons, at least not to any important extent.

In conditions of scarcity, increased bids from unsatisfied buyers would send up prices without greatly increasing supplies. Marginal cost rises steeply for a small increase in output, and price is again equal to marginal cost, which includes costs of diverting supplies from normal channels. The richer members of the community may enjoy their normal consumption, but the poorer members will get little or none.

Collective Demands and Supplies

The State does not confine its economic activities to the definition of the framework inside which production takes place. We have seen that the result of this form of activity is a modification of demand and supply schedules, and the further type of activity we now have in mind takes the form of direct instead of indirect intervention in these schedules.

The State has always regarded certain productive activities as so important that it alone should undertake them, and through time the scope of such activities has steadily widened. The provision of external defence, internal order, health and education services and the like means that the State must secure command of factors of production which would otherwise be devoted to other purposes. It can do this by requisi-

[1] Marginal cost in this context will be the extra cost to the Government of increasing the national supply of meat. The Government will presumably buy supplies from various sources, at home and abroad, and while it will pay the same price to all sellers of a particular grade in any one country, it need not pay the same price in every country. Marginal cost will therefore be the price it pays to the dearest group of sellers.

tion in the way the personnel, and formerly the equipment, of armed forces have been provided, but the method of conscription of men and requisition of property is, in general, less efficient than the method of acquisition through the market. So the State is a considerable purchaser of factors of production, and it pays for the factors it needs from funds raised by taxation. Sums which would be spent by individuals as *their individual* tastes might direct are spent by the State in accordance with collective desires. Some of the collective demands expressed by the State would exist as individual demands if the State made no provision, but it may be that, as with defence, water supplies, main drainage and highways, individual provision is either difficult or even impossible. In each case in a democratic State, at least, collective provision is adopted where it is regarded as being superior to individual provision.

Taxation as a Fee Payment

In certain cases—in general those first selected for collective provision—it is not possible to compute the benefit derived by the individual from the State-provided service, because one cannot say how much of that service the individual consumes. It is impossible to say what is the value of the police services, the main drainage or the services of highways one consumes in a year, so that a charge based on use of service is impossible. Consequently, the attempt is made to distribute the cost of public services among those who use them according to some arbitrary criterion of the benefit enjoyed. This arbitrary criterion may be the amount of land owned or occupied, the number and size of houses possessed, the quantities of certain types of goods purchased, or the amount of the individual's money income. The alternative is to levy a price for the use of facilities, as when a toll for the use of a road or a bridge is demanded each time use is made of it. The choice between one method and the other depends largely on differences in individual tastes. Differences in individual consumption of milk and bread would lead State dairies and bakers to make a charge per unit of purchase, but our consumption of road services is not confined to the amount we walk or ride upon them.

Similarly, the benefits of education services accrue, not merely to those who receive free education for their children,

but all share in some degree in the greater productivity of a literate society as compared with an illiterate one. The forces behind popular education in the nineteenth century were the industrial and commercial classes who wanted labour which could read instructions, write their letters and keep their books. Health services benefit, not only those who receive free medical attention, but all who are freer from risk of infection or whose employees lose less time from illness because of public medical services. They facilitate the operation of the social machine on which each is dependent for his livelihood.

It is a logical step from this view of the matter to the conclusion that the amount of tax paid by an individual should depend on the size of income enjoyed by the individual in question. It has been argued that the greater the income an individual enjoys, the greater the use he makes of the general services of the State in maintaining the organisation of production which makes that income possible, so that, although a precise computation of the value of those services cannot be made, an allocation of the cost of the services on the basis of income received will be roughly proportional to the use made of the services, and will therefore be socially equitable. The State might therefore raise revenue by a tax system for general services which cannot be priced, and charge fees in cases where individual consumption can be valued.

Relation of Tax to Income

This conclusion still leaves open the question of whether the proportion of income paid in general taxation should increase, decrease or remain constant as income increases; that is, whether taxation should be progressive, regressive or proportional. The utility economists justified progressive taxation on the ground that the marginal utility of income decreases as income increases, so that the amount of satisfaction derived by an individual from the $(N+1)$th unit of income is always less than the satisfaction derived from the Nth unit. We have earlier denied the possibility of assessing the amount of satisfaction derived from the whole or part of income, but this defence of progressive taxation involves a step which is still more dubious. For it to be true, it is necessary to assume that the satisfactions derived by any two individuals from the Nth

units of income they enjoy are the same. Then, if it is considered equitable to deprive every individual by taxation of one-tenth of the satisfaction he could enjoy by spending the whole of his income, it will be necessary to take a larger proportion of the income of a richer man than of a poorer man.

Suppose that *A* has an income of £500 and *B* an income of £1,000, and that *A* pays a tax of £50, then if *B*'s sacrifice is to be proportionately equal to *A*'s, it will be necessary to make *B* pay more than £100. *A*, when he is taxed, will lose the satisfactions which those units of his income numbered 451–500 could bring him, but the satisfaction he loses through the subtraction of the 500th unit, i.e. the least important to him, is greater than *B* loses by the subtraction of the 901st, the most important unit he loses. Therefore, the satisfaction which *B* loses by paying tax of £100 is less than twice that which *A* loses through paying £50, and for sacrifices to be the same for both, *B* must pay more than £100. The argument collapses if we cannot admit interpersonal comparisons of utility.

If we take the approach of regarding taxation as a payment for the general services rendered by the State, we have to assume that consumption of such services increases with income but at a greater rate, a probable but quite unprovable conclusion. The problem is indeed quite insoluble, but fortunately it is not a real problem at all. The fundamental principles which underlie the distribution of taxation are indeed not economic at all but political. Taxation policies are designed to achieve certain ends selected by the class which exercises political power, and these choices have to be accepted by the economist, just as preferences for white instead of brown bread, or for beer instead of wine, have to be accepted. The economist can only work out their implications. Where a wealthy aristocracy have a monopoly of political power, taxation will be regressive; where the masses rule, taxation will be progressive.

The imposition of taxation must inevitably modify the fundamental economic data; demands for goods and services; willingness to work, to save and to take risks; the choice between letting land for rent and keeping it for private enjoyment. In the nineteenth century, when the political philosophy of the time declared that money fructifies in the pockets of individuals but stagnates in the coffers of the State, the task

of the economist who studied public finance was to show how to raise the funds to pay for certain services which it was considered desirable the State should provide, with the minimum interference in the general conditions of demand and supply. Under other regimes taxation may have the deliberate purpose of diverting income from one group of the community to another. History provides numerous examples where a peasant population has been heavily taxed to build pyramids or cathedrals or to satisfy the lusts of a monarch for luxury or military adventure. On the other hand, a socialist regime may pursue a policy of " soaking the rich," that is, of taxing large incomes heavily to provide free services which will be consumed mainly by lower income groups.

Taxation as an Instrument of Economic Policy

We have now stepped into an entirely distinct field, where taxes are imposed, not merely to raise money to cover certain expenditures, but because of their effects on the distribution of incomes and on supplies and demands. Thus during war the central task of economic policy is to divert the maximum quantity of factors of production away from their normal occupations to the production of munitions. At the same time the population must be induced to put forth the maximum effort, and it may not be easy to pursue these policies simultaneously. There may be cases where the production of certain commodities requires factors which are particularly necessary for munitions production. If it is not desired to prohibit manufacture entirely, a purchase tax may be imposed to reduce demand. Such a tax is useful when it is desired to reduce output over a wide field of commodities where rationing is impracticable. The reduction in consumption may be a much more important objective than the amount of tax collected.

It might be possible to cut down consumption to the requisite extent and also provide the funds to pay for the war by imposing an income-tax at such a rate that it left only the means of purchasing bare essentials. Such a policy would, however, be calculated to reduce incentive to work and to lead to lower output. It may be advisable, therefore, to allow people to receive incomes which, after taxation has been met, will leave a margin over and above that necessary to maintain physical

efficiency, and by savings campaigns to induce people to lend the balance to the Government.

At the same time the Government will wish to see that all receive a certain minimum of primary necessities and to ensure that the rising cost of such necessities does not lead to a general rise in wages and prices which might disrupt the whole system. It may therefore provide a ration of such commodities at a fixed price, subsidising the sale of these goods if for any cause their prices tended to rise. In this way increases in the cost of living due to external causes, such as a rise in the price of imported food-stuffs or materials, and consequent demands for wage increases, will be avoided.

Changes in the form of taxation may be used to counter the changes in income and employment which are associated with the trade cycle. If it is wished to increase propensity to consume, a removal of purchase taxes or a redistribution of the tax burden in favour of the lower income groups and to the disadvantage of the upper income groups may have this effect. If investment needs encouragement, a reduction in the rate of tax on earned income or an increase in allowances made for depreciation may have this effect. We cannot say that these effects will inevitably follow; much depends on particular circumstances, and in the wrong circumstances a reduction in tax burden may lead to an attempt to acquire larger cash balances.

Again, a reduction in taxes on earned incomes may provide an incentive to greater effort and increase willingness to work overtime. This is particularly the case where the earned income-tax rate is steeply progressive, so that the first £*x* of income, say, are free of tax, the next £*y* taxed at a low rate, the next £*z* at a higher rate and so on. If each step in the tax scale is steep, it will mean that, for people who are slightly above or below it, the marginal rate of tax is much greater than the average rate. So, if the rate on the first £100 is nil, the rate on the second 2*s*. 6*d*., on the third 5*s*., and on everything after that 10*s*., a man earning £200 will pay an average rate of 1*s*. 3*d*., but if he works overtime he will pay 5*s*. on each additional pound and at the next step the average rate is half a crown, but the marginal rate ten shillings. If the second tax step is put at £250 instead of at £200, people whose incomes are round the latter figure may feel it more worth their while to work over-

time or to speed up their rate of working on piece-rates, or to cut down absenteeism and increase their earnings. If one of these tax steps happens to correspond with the income level of a large group of people, the effect may be important.

The State and Control of the Quantity of Money

Control of the currency has generally been undertaken by the State, because no lesser authority could be trusted to see that coins were of the weight and fineness indicated by their face value. Debasement of coinage could be so easily effected that the power of the State with all the sanctions it could bring to bear against the forger was considered essential to the maintenance of its quality. The crown itself was guilty of debasement on occasions, but other authorities were considered even more fallible.

Notes, on the other hand, were commonly issued by private organisations under State authority, but the temptation to over-issue was always strong, and more recently note issue also has tended to become a State monopoly. In Britain, even when the State issued no notes, it for long controlled the quantity issued in the interest of maintaining its value in terms of gold. The quantity of them had to be regulated, so that notes were always a close substitute for gold. Thus, the various Bank Acts laid the obligation on the Bank of England to redeem its notes in gold on demand, and to hold an equivalent quantity of gold for every note issued beyond an amount defined by statute. Similar limitations were imposed on commercial banks who were allowed to issue notes, but in their case they could convert into gold or Bank of England notes, and their reserve fund might also have this dual composition. Thus, people who preferred notes to gold could always exchange on the basis of one five-pound note for five sovereigns, and those who preferred gold could always get five sovereigns for a note.

As we have seen, the total quantity of currency, gold and notes was regulated in such a manner as to check any tendency for gold to flow in or out of the country. If gold was flowing into a country, more metal was coined and more notes issued, so that prices tended to rise, checking exports and encouraging imports, and thus diminishing the favourable balance of payment for all items other than gold. A contrary course was

pursued if gold were tending to flow out of the country. During the second quarter of this century the objective of monetary policy has changed from that of maintaining the external value of the currency unit to that of maintaining the volume of employment or the national income. Indeed, in efforts to maintain employment, Governments have deliberately devalued their currencies in terms of those of other countries in the hope of stimulating exports while leaving internal prices and wages unchanged.

Regulation of External Trade

It has been a common practice in the past for the State to impose restrictions on trade crossing its frontiers; indeed, in Britain, where there was more enthusiasm for free trade than elsewhere, it lasted some three-quarters of a century, and at no time was there complete absence of restriction. There have been prohibitions on imports or on exports, taxes on both types of trade, and quantitative regulation by means of quotas. The export of English wool was once prohibited to hinder the growth of rival industries, and for the same reason the export of machinery has been banned. The import of cotton goods has been banned to preserve the home market for the native woollen industry. Tariffs have been placed on the import of food-stuffs for a variety of reasons: to ensure an indigenous supply in time of war; to maintain the agricultural population at a certain level for pseudo-sociological reasons or to secure a supply of soldiers; or merely because "wheat is sacred to Frenchmen." Industries have been fostered behind tariff walls for reasons which can invariably be interpreted as efforts by interested parties to conceal the real purpose, the creation for themselves of an investment opportunity which would not otherwise exist.

Tariffs

Popular views on tariffs are largely the result of erroneous analogies between the economy of a firm and the economy of a community. In a firm the greater its sales in relation to its purchases, the greater are net receipts which constitute its disposable income. For the external trade of a community there is no quantity analogous to net receipts. The total payment

for its exports of goods and services is its disposable external income, and if this income is to contribute to the consumption income of the community, it must be brought home in the form of goods or services purchased. Further, if it is to contribute to investment inside the country, it must also be brought home in the form of imports. Unless it is brought home in one of these two ways, it must remain abroad in the form of external investment and then the same problem arises over the income from the investments, and a community is hardly likely to continue indefinitely to acquire investments from which it draws no income.

Looking at the problem from the point of view of the country against whose goods the barrier is raised, if its exports to the protectionist country are reduced, or their price lowered, the consequent reduction in its external income will force it to reduce either its purchases abroad or its external investments. In the second case the reduction in external purchases will be shifted to the countries where the investment has been reduced, so that the final result will be the same. Each country which experiences a reduction in demand for its exports will be obliged to reduce its demand for imports, and the effect will in the end be felt by the country which imposed the tariff in the first place.

There is nothing more profound in this than the elementary fact that the two sides of a balance-sheet must balance when all items are taken into account. There must be a balancing of payments between each country and the rest of the world taken together. The total of payments received by the Government and residents in a particular area must equal the total of payments made by the same people to persons and institutions outside that area. The payments on current income account may be in respect of goods exchanged, services rendered, property hired or interest on money lent. If there is a net payment on income account in either direction, it can exist only because some person or institution in the area which has a surplus lends to a person or institution in the area with a deficit. This is true of the balance of payments between one individual and the rest of humanity, between a village and the rest of the country, or between the United Kingdom and the rest of the world. No individual or community can receive

payments in excess of the value of the goods and services it sells, except in so far as it can secure loans or gifts.

If a country wished to enhance its income by manipulating volumes of imports and exports, it should contrive to decrease its exports and increase its imports, and finance the difference by an ever-increasing stream of perpetual loans; in other words, to create for itself a spendthrift's paradise. The contrary policy may accord with orthodox principles of financial rectitude, but if the excess of exports is achieved by means of a tariff plus external lending, both total external disposable income and total consumption would appear to be less than they would be without the tariff.

Taxing the Foreigner

Must we then conclude that those who wish, by means of a tariff, to tax the foreigner are entirely mistaken? It is indeed possible to construct a set of circumstances in which such a result would be achieved. If, because it has alternative sources of supply, country *A* has an elastic demand for the products of country *B* whose supply to *A* is very inelastic, and it imposes an import duty on goods from *B* only, then it can force producers in *B* to pay most of the tax, but it will still be open to citizens of *B* to reduce their purchases of *A*'s goods. Therefore, it is necessary also that *B*'s demand for *A*'s goods should be very inelastic. Even so, the citizens of *B* would have to cut down their external purchases somewhere, and so they might buy less from *C* and *D*, who in turn would buy less from *A*. This would be obviated if the proceeds of the tax collected by *A* were used to buy the goods of *C* and *D*, which *B* would otherwise have bought. The circumstances are extremely improbable, and in the unlikely event of their occurrence, the imposition of a discriminatory tariff against the products of one country would evoke vigorous retaliation, but only if they were fulfilled and no retaliation took place would it be possible to tax the foreigner.

Do Tariffs Create Employment?

It is argued that a tariff may be used to export unemployment. By shutting out imports it is possible to create jobs making the things formerly imported. Similarly, we can all

create lots of work for ourselves by cutting our own hair, making our own clothes, and performing all the tasks we normally pay other people to perform. But shall we be any better off if we do it? We certainly shall not if it means that we have to do less work of a kind for which we are paid rates higher than we pay to those who perform services for us. Yet this is what happens when production is moved, from a location where costs are lower to one where costs are higher, by means of a tariff. So long as there is employment for all factors of production in a country, these factors will produce a larger income, if they are allowed to remain in industries in which they have settled because of their natural advantages for those occupations rather than if they are diverted to industries which cannot establish themselves in the environment without assistance. If there is a long-run tendency to under-employment, it is possible that a tariff might increase both employment and income, but it is a crude tool to use for the purpose.

The disadvantage of using a tariff as an instrument of economic policy is that any positive results it may achieve can be readily seen, but its negative results cannot. The number of people drawn into employment can be counted, but the number thrown out of employment because of the fall in demand for exports cannot be demonstrated. The channels of international trade are not sufficiently well known to allow the connection between individual imports and exports to be traced.

Political Systems and Economic Analysis

This discussion of some of the economic problems of state policy has served to amplify what was said in the first few pages of this book about the nature of economic science. To continue the quotation from J. M. Keynes to which we then referred, the Theory of Economics is a "method rather than a doctrine, an apparatus of the mind, a technique of thinking, which helps the possessor to draw correct conclusions." Analysis is independent of the data on which it has to work. The conclusions it reaches depend on the data on which it works.

As in every science, the data is provided by observation, and consists, on the one hand, of a set of objective facts concerning the physical world and, on the other, of a set of subjective facts

concerning the people involved, in their capacities as producers or consumers. It is the latter group of considerations which is affected by the characteristics of the political system. The willingness of a given group of people to supply factors of production will be different according to whether the decisions to supply are collective or individual decisions: demands will differ according to whether they are expressed by individuals through the market or collectively by public authorities.

The policy appropriate to a particular occurrence in one political setting may be entirely inappropriate in another political setting.

A further complication arises from the fact that it is rare that an economic problem has a unique solution. The choice must be made between the alternatives in the light of the whole of the effects of the particular action taken.

The effects will commonly involve ethical, æsthetic and political as well as economic considerations, so that economic analysis will never tell us which of the possible alternatives we should choose. Used skilfully, it will set out the alternative lines of action and the consequences of each. Its *raison d'être* is that it makes intelligent choice in social problems possible.

Further Reading

DALTON, H.: *Principles of Public Finance.*
HICKS, J. R.: *Social Framework.*
HICKS, U. K.: *Public Finance.*
HICKS, U. K.: *Finance of British Government,* 1920–36.
MEADE, J. E : *Planning and the Price Mechanism.*

More Advanced Reading

DE VITI DE MARCO: *First Principles of Public Finance.*

INDEX

लाल बहादुर शास्त्री राष्ट्रीय प्रशासन अकादमी, पुस्तकालय
Lal Bahadur Shastri National Academy of Administration Library
MUSSOORIE/मसूरी
अवाप्ति सं० /Acc. No. 5870

कृपया इस पुस्तक को निम्नलिखित दिनांक या उससे पहले वापस कर दें।

Please return this book on or before the date last stamped below.

दिनांक Due Date	उधारकर्ता की संख्या Borrower's No.	दिनांक Due Date	उधारकर्ता की संख्या Borrower's No.

330.15
Eas
5870
LBSNAA

330.15
Eas

अवाप्ति संख्या
Acc. No. 5870

वर्ग संख्या
Class No. ______

पुस्तक संख्या
Book No. ______

लेखक
Author Eastham, J. K.

शीर्षक
Title Introduction

330.15
Eas.

LIBRARY
LAL BAHADUR SHASTRI
National Academy of Administration
MUSSOORIE

Accession No. 5870

1. Books are issued for 15 days only but may have to be recalled earlier if urgently required.
2. An over-due charge of 25 Paise per day per volume will be charged.
3. Books may be renewed on request, at the discretion of the Librarian.
4. Periodicals, Rare and Reference books may not be issued and may be consulted only in the Library.
5. Books lost, defaced or injured in any